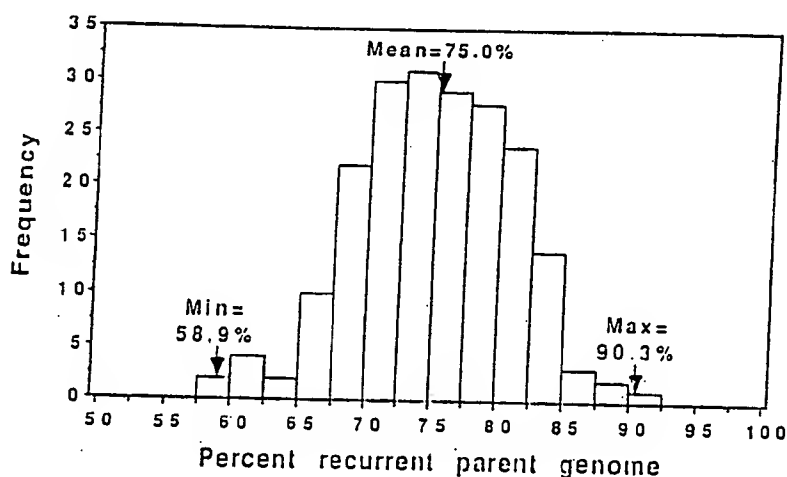




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁵ : C12Q 1/68, A01H 1/04</p>	<p>A1</p>	<p>(11) International Publication Number: WO 90/04651 (43) International Publication Date: 3 May 1990 (03.05.90)</p>
<p>(21) International Application Number: PCT/US89/04688 (22) International Filing Date: 19 October 1989 (19.10.89) (30) Priority data: 259,998 19 October 1988 (19.10.88) US (71) Applicants: WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US). CORNELL RESEARCH FOUNDATION, INC. [US/US]; East Hill Plaza, Ithaca, NY 14850 (US). (72) Inventors: LANDER, Eric, S. ; 151 Bishop Richard Allen Drive, Cambridge, MA 02139 (US). PATERSON, Andrew, H. ; 1026 Ellis Hollow Road, Apt. 32, Ithaca, NY 14853 (US). TANKSLEY, Steven, D. ; 215 Connecticut Hill Road, Newfield, NY 14867 (US).</p>		<p>(74) Agents: GRANAHAN, Patricia et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US). (81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>

(54) Title: MAPPING QUANTITATIVE TRAITS USING GENETIC MARKERS



(57) Abstract

A systematic and accurate method for mapping or locating the genomic regions or subregions containing polygenic factors which control a quantitatively inherited trait or traits of interest. The present method, which can be applied to higher plants and animals, makes it possible to determine with a high degree of accuracy that a quantitative trait loci (QTL) lies within a genomic region or subregion bounded by selected genetic markers.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MR	Mauritania
BE	Belgium	GA	Gabon	MW	Malawi
BF	Burkina Faso	GB	United Kingdom	NL	Netherlands
BG	Bulgaria	HU	Hungary	NO	Norway
BJ	Benin	IT	Italy	RO	Romania
BR	Brazil	JP	Japan	SD	Sudan
CA	Canada	KP	Democratic People's Republic of Korea	SE	Sweden
CF	Central African Republic	KR	Republic of Korea	SN	Senegal
CG	Congo	LI	Liechtenstein	SU	Soviet Union
CH	Switzerland	LK	Sri Lanka	TD	Chad
CM	Cameroon	LU	Luxembourg	TG	Togo
DE	Germany, Federal Republic of	MC	Monaco	US	United States of America
DK	Denmark				

-1-

MAPPING QUANTITATIVE TRAITS USING GENETIC MARKERSDescriptionBackground

05 Although some important characteristics of
agricultural crops and animals are determined by genetic
loci which have major effects on phenotype, most traits
which are economically valuable are quantitative in
nature. In the case of a quantitatively inherited trait,
phenotypic variation in that trait shows continuous
10 quantitative differences among individual plants or
animals and there is a normal distribution of phenotypic
values for the trait in a given population. The
continuous variation is seen because of the fact that the
collective effect of numerous genetic loci, each having a
15 small effect (making a small quantitative contribution
to), determines the phenotypic trait. Traits which
exhibit such genetic variation are often referred to as
polygenic traits because genetic variation at a large
number of genetic loci affects phenotypic expression.
20 Loci which contribute to polygenic traits are often
referred to as "minor genes"; "major genes" are those in
which variation results in substantial effects on
phenotype and which exhibit a Mendelian pattern of
inheritance. Beckmann, J.S. and M. Soller, Oxford
25 Surveys of Plant Molecular and Cellular Biology,
3:196-250 (1986).

-2-

The conflict between the Mendelian theory of particulate inheritance and the observation that most traits in nature exhibit continuous variation was resolved in the early 1900s by the concept that

05 quantitative traits can result from segregation of multiple genes, modified by environmental effects. Mendel, G., Verh. des Naturf. Vereines in Brunn, 4 (1866); Johannsen, W. Elemente der exakten Erbilichkeitsllehre (Fischer, Jena, 1909); Nilsson-Ehle,

10 H. Kreuzunguntersuchungen an Hafer und Weizen (Lund, 1909); East, E.M., Genetics, 1:164-176 (1915); Wrigth, S. Evolution and the Genetic of Populations (Univ. of Chicago, Chicago, 1968). Although it has been shown that polygenes exhibit a Mendelian pattern of inheritance, it

15 has not been possible to separate the effect of different loci acting in this way. Pioneering experiments have shown that linkage to such quantitative trait loci (QTLs) could occasionally be detected. However, accurate and systematic mapping of QTLs has not been possible because

20 the inheritance of an entire genome could not be studied with genetic markers. Sax, K., Genetics, 8:552-556 (1923); Thoday, J.M., Nature, 191:368-369 (1961); Tanksley, S.D. et al., Heredity, 49:11-25 (1982); Edwards, M.D. et al., Genetics, 116:113-125 (1987). The

25 use of restriction fragment length polymorphisms (RFLPs) has made such investigations possible, at least in principle. Botstein, D. et al., American Journal of Human Genetics, 32:314-331 (1980).

DNA polymorphisms, which are differences in the

30 nucleotide sequence of a region of DNA, are of several types. For example, variation in the nucleotide sequence of DNA which is the result of a point mutation can, in turn, result in gain or loss of a restriction site for a

- 3 -

particular restriction endonuclease. Changes in DNA which involve larger regions (deletions, additions, inversions, translocations) change the relative distribution of restriction sites for several restriction endonucleases. In both cases, endonuclease digestion of the region affected by the mutation produces DNA restriction fragments which differ in size distribution from fragments similarly obtained from an unaffected individual (i.e., one in whom such a mutation or alteration is not present). The result is an RFLP, which, in the case of a point mutation, is generated only by enzymes whose recognition sites include the mutation and, in the case of changes in larger regions, is generated by a number of enzymes.

RFLPs have been shown to be stably inherited and, in the case of nuclear RFLPs, to express codominance and generally lack obvious phenotypic effects. In addition, RFLPs often exhibit multiple alleles. Assigning locations of RFLPs on a chromosome map can be carried out, using art-recognized techniques. Beckmann, J.S. and M. Soller, Oxford Surveys of Plant Molecular and Cellular Biology, 3:196-250 (1986).

Manipulation of QTL has historically been a major limitation of genetic engineering and classical breeding. Systematic and accurate mapping of QTLs has not been possible because of the difficulty in arranging crosses with genetic markers densely spaced throughout an entire genome. RFLP techniques make it possible to try to identify and manipulate QTLs, by, for example, determining RFLP-QTL linkage, followed by mapping and evaluating QTL effects; directly identifying allelic variation at a QTL; and using insertional mutagenesis to identify and clone QTL. It would be very beneficial if a

-4-

method by which QTL could be identified, accurately mapped and introduced, as needed, into plants and animals, particularly for development of new strains which exhibit desirable features.

05 Summary of the Invention

The present invention relates to a systematic and accurate method for mapping or locating precisely the genomic regions containing polygenic factors controlling a quantitatively inherited trait or traits of interest.

10 Described herein is the first method by which accurate mapping of a QTL to an interval can be carried out. That is, unlike previously-available methods of determining QTLs, the present method makes it possible to determine with a high degree of accuracy that a QTL lies within an

15 interval (e.g., within a region of DNA bounded by two selected markers or a subregion thereof). The accuracy with which mapping can be carried out by the method of the present invention is particularly valuable in that it provides precision in at least two applications: gene

20 cloning and gene transfer. For example, in gene cloning, the present method makes it possible to delimit the region in which it is highly likely that DNA encoding a quantitative trait of interest occurs, thus maximizing the likelihood that the DNA of interest is, in fact

25 isolated and minimizing the effort which must be expended in its isolation. Similarly, the present method is also very valuable in those instances in which transfer of a gene or gene portion from one plant or animal (e.g., an unagricultural product or wild type) to another plant or

30 animal (e.g., an agricultural product) is to be carried out. As in the case of gene cloning, in gene transfer, tight intervals increase the likelihood that the desired

- 5 -

gene will be transferred and decrease the effort (and concomitant time and expense) necessary to transfer the gene or gene portion.

05 The method of the present invention makes use of
genetic markers to map/locate QTL, to estimate or predict
their phenotypic effects and to greatly reduce the number
of progeny which must be scored with the DNA markers. In
the case described herein, RFLPs, isozymes and two
specific genes were the genetic markers used. However,
10 any genetic markers, such as any DNA polymorphisms or DNA
sequence differences that can be detected, codominant
protein polymorphisms or a combination of codominant
genetic markers can be used for the same purpose. Any
set of scorable genetic markers (codominant or recessive)
15 which cover most of the genome of the plant or animal
being assessed can be used for this purpose.

As a result of following an entire genetic map of
RFLPs in a cross of two strains which differ for one or
more quantitatively inherited traits, it has been
20 possible to develop analytical methods useful for
systematic and accurate mapping of these genes and, then,
by following the inheritance of nearby markers, for
breeding them into new strains. The availability of
complete RFLP linkage maps, such as described herein, has
25 made is possible to dissect quantitative traits into
discrete genetic factors. Subsequently, QTLs can be
mapped and isogenic lines can be constructed, so as to
differ only in the region of the QTL, by using the RFLPs
to select for the desired region and against the
30 remainder of the genome. Such isogenic lines can be
used, in conjunction with the fundamental tools of
genetic and molecular biology for the study of a trait,
including testing of complementation and epistasis;

-6-

characterization of physiological and biochemical differences between isogenic lines; isolation of additional alleles via mutagenesis or further selective breeding (at least in favorable systems); and, eventually, molecular cloning of the genes underlying quantitative inheritance. The method of the present invention has broad application to breeding of plants and animals for agriculturally valuable traits, particularly because it allows for deterministic breeding.

Systematic genetic dissection of quantitative traits using complete RFLP linkage maps is valuable in a broad range of biological endeavours. For example, agricultural traits, such as resistance to diseases and pests, tolerance to drought, heat, cold, and other adverse conditions, and nutritional value can be mapped and introgressed into domestic strains from exotic relatives (Rick, C.M., Genes: Enzymes and Populations, Plenum Press, pp. 255-268 (1973); Harlan, J.R., Crop Science, 16:329-333 (1976)). Aspects of mammalian physiology such as hypertension, atherosclerosis, diabetes, predispositions to cancer and teratomas, alcohol sensitivity, drug sensitivities and some behaviours can be investigated in animal strains differing widely for these traits (Tanase, H. et al., Japanese Circulation Journal, 34:1197-1212 (1970); DeJong, W., Handbook of Hypertension, Vol. 4, Elsevier (1984); Paigen, B.A. et al., Atherosclerosis, 57:65-73 (1985); Prochazka, M. et al., Science, 237:286-289 (1987); Heston, W.E., J. Natl. Cancer Inst., 3:79-82 (1942); Kalter, H., Genetics, 39:185-196 (1954); Malkinson, A.M. & D.S. Beer, J. Natl. Cancer Inst., 70:931-936 (1983); Shire, J.G.M., In: Genetic and Environmental Influences on Behavior, pp. 194-205 (1968);

-7-

Stewart, J. & R.C. Elston, Genetics, 73:675-693 (1973);
Festing, M.F.W., Inbred Strains in Biomedical Research,
Oxford University Press, Oxford (1979)). Evolutionary
questions about speciation can be elucidated by
05 determining the number and nature of the genes involved
in reproductive barriers (Coyne, J.A. & B. Charlesworth,
Heredity, 57:243-246 (1986)). An example of such genetic
dissection is described herein: In an interspecific
cross in tomato, QTLs affecting fruit weight,
10 concentration of soluble solids and fruit pH are mapped
to within about 20-30 centiMorgans (cM, which = 1%
recombination about 10^6 - 10^7 bp) by means of a complete
RFLP linkage map.

Brief Description of the Drawings

15 Figure 1 is the frequency distribution for fruit
mass, soluble-solids concentration ($^{\circ}$ Brix, a standard
refractometric measure primarily detecting reducing
sugars, but also affected by other soluble constituents;
 1° Brix is approximately 1% w/w) and pH in the E parental
20 strain and in the backcross (BC) progeny.

Figure 2 is the distribution of recurrent parent (E)
genotype in the 237 backcross progeny, estimated on the
basis of the marker genotypes and their relative
distances.

25 Figure 3 are QTL likelihood maps indicating LOD
scores for fruit weight (solid lines and bars), soluble
solids concentration (dotted lines and bars) and pH
(hatched lines and bars), throughout the 862 cM spanned
by the 70 genetic markers. The RFLP linkage map used in
30 the analysis is presented along the abscissa, in Kosambi
cM.

- 8 -

Figure 4 is a schematic drawing of phenotypic distributions in the A and B parental, F1 hybrid and B1 backcross populations.

Figure 5 are graphic representations of LOD scores
05 for a hypothetical quantitative trait, based on simulated data for 250 backcross progeny in an organism with 12 chromosomes of 100 cM each.

Figure 6 is a graphic representation of LOD scores for a chromosome containing two QTLs.

10 Figure 7 shows the appropriate LOD threshold so that the chance of a false positive occurring anywhere in the genome is at most 5%, as a function of genome size and density of RFLPs scored.

Figure 8 shows that progeny having phenotypes
15 exceeding mean by $\geq L$ standard deviations make up a proportion $Q(L)$ of population, but account for a proportion $S(L)$ of the total LOD score for the progeny.

Figure 9 shows that if only individuals having
20 phenotypes exceeding mean by $\geq L$ standard deviations, the number of progeny genotyped may be decreased by a factor of $g(L)$ if the number of progeny grown and phenotyped is increased by a factor of $h(L)$.

Figure 10 shows the number of backcross progeny that must be genotyped to map a QTL, based on the fraction of
25 the backcross variance explained by the segregation of the QTL. In the traditional approach, all progeny are genotyped and single markers are analyzed. In the method of the present invention, only progeny with 5% most extreme phenotypes are genotyped and interval mapping is
30 used to analyze the data.

Figure 11 shows the number of backcross progeny that must be genotypes to map a QTL, based on the difference D between the strains (measured in environmental standard

-9-

deviations) and the number K of effective factors. In the traditional approach (Panel A), all progeny are genotyped and single markers are analyzed. In the method of the present invention (Panel B), only progeny with 5% most extreme phenotypes are genotyped and interval mapping is used to analyze the data.

Detailed Description of the Invention

The present invention is based on a method of resolving quantitative traits into discrete Mendelian factors, using genetic markers. In the work described herein, a complete RFLP linkage map was available and RFLPs were the genetic markers used. However, any genetic markers which will generally be codominant markers, such as DNA polymorphisms, isozymes or other codominant protein polymorphisms or a combination of such markers, can be used. Described herein is the first use of such a complete RFLP linkage map to resolve quantitative traits into discrete Mendelian factors, in an interspecies backcross of a higher plant. In the instant case, a series of QTLs controlling selected traits in the higher plant were mapped, using the new analytical methods described herein. The methods described herein, which can be applied in a similar manner to mapping QTLs in other plants and in animals, have made it possible, for the first time, to map QTLs to DNA intervals with a high degree of accuracy, thus maximizing the likelihood that a QTL of interest lies within an interval defined by two selected (defined) markers. This approach is broadly applicable to the genetic dissection of quantitative inheritance of physiological, morphological and behavioral traits in any higher plant or animal.

-10-

The method of interval mapping of the present invention and its application to a higher plant are described in detail herein. The method of interval mapping of the present invention can also be used to

05 locate precisely the genomic regions which contain polygenic factors controlling a quantitatively inherited trait or traits. Briefly, the subject method entails five steps or procedures: 1) choosing of a pair of

10 interfertile strains (i.e., two types or varieties of a plant or animal which differ as to a trait of interest), which will serve as the parent strains in the initial cross (and one of which may serve as the recurrent parent in subsequent crosses if backcrosses are employed); 2)

15 constructing a genetic linkage map (using RFLPs, isozymes and/or other codominant markers), if an adequate map is not already available; 3) arranging one or more back-crosses or intercrosses, using as the recurrent parent the strain or type of plant or animal in which the transferred gene (or genes) is to function; 4) scoring

20 progeny of the back-crosses or intercrosses for the trait or traits of interest and for the genetic markers comprising the linkage map; and 5) applying an algorithm designed to maximize the likelihood of a specific/select-

25 ed function based on the data obtained in (A).

25 An example of application of the interval mapping method of the present invention to locate genomic regions containing QTLs of interest in tomato plants is presented in the following sections.

30 The following is a description of: a) the parent plants used in the interspecies backcross; b) assessment of backcross progeny; c) construction of a genetic linkage map; d) interval mapping of QTLs; e) a summary or

-11-

overview of key considerations; and f) possible applications of QTL mapping based on the resulting data.

a. Parent plants

Described below is resolution of quantitative traits, using a complete RFLP linkage map, in an interspecific back-cross of two types of tomato plants: the domestic tomato Lycopersicon esculentum (L. esculentum) cv. UC82B (denoted E) and a wild South American green-fruited tomato L. chmielewskii (L. chmielewskii) accession LA1028 (denoted CL). Chmielewski, T., Genet. Polon., 9:97-124 (1968).

These strains have very different fruit masses (E approximately 65 g; CL approximately 5 g) and concentrations of soluble solids (E approximately 5%; CL approximately 10%). These are traits of agricultural importance because they jointly determine the yield of tomato paste. Rick, C.M., Hilgardia, 42:493-510 (1974). In addition, the strains are known to be polymorphic for genes affecting fruit pH, which is important for the optimal preservation of tomato products; the difference in pH between parental strains is, however, small. Tanksley, S.D. and J. Hewitt, J. Theor. Appl. Genet., 75:811-823 (1988).

b. Back-cross progeny assessed

A total of 237 back-cross plants, with E as the recurrent parent, were grown in the field at Davis, California. Between five and 20 fruit from each plant were assayed for fruit mass, soluble-solids concentration ($^{\circ}$ Brix) and pH, each of which showed continuous variation (Figure 1). Tanksley, S.D. and J. Hewitt, J. Theor.

-12-

Appl. Genet., 75:811-823 (1988). °Brix is a standard refractometric measure which primarily detects reducing sugars; it is also affected by other soluble constituents. 1°Brix is approximately 1% w/w. Example 1
05 is a detailed description of the back-cross and assessment of backcross progeny.

c. Construction of a genetic linkage map

A genetic linkage map of tomato with more than 300
10 RFLPs and 20 isozyme markers had previously been constructed, by analyzing 46 F₂ individuals derived from L. esculentum cv. VF36 x L. pennellii accession LA716 (E x P). Tanksley, S.D. et al., R. Proc. 18th Stadler Genet. Symp., In Press. The map is essentially complete:
15 it has linkage groups covering all 12 tomato chromosomes, with an average spacing of 5 cM between markers (1 cM is the distance along the chromosome which gives a recombination frequency of one per cent). For QTL mapping, a selected subset of markers spaced at
20 approximately 20 cM intervals and displaying polymorphism between the E and CL strains was used. These included 63 RFLPs and five isozyme markers. In addition, the E and CL strains differ in two easily-scored, simply-inherited morphological traits: determinancy (described below) and
25 uniform ripening, controlled by the sp and u genes, respectively. Although a few distal regions did not contain appropriate markers, it is estimated that about 95% of the tomato genome was detectably linked to the markers used.

30 These 70 genetic markers were scored for each of the 237 E x CL back-cross progeny (as described by Tanksley and Hewitt) and a linkage map was constructed de novo using MAPMAKER. Tanksley, S.D. and J. Hewitt, J. Theor.

-13-

Appl. Genet., 75:811-823 (1988); Lander, E.S. et al., L. Genomics, 1:174-181 (1987). The map covers all 12 chromosomes with an average spacing of 14.3 cM. Although the linear order of markers inferred from the E x CL cross essentially agreed with that inferred from the E x P cross, differences were noted (See Example 1). Genetic distances differed markedly in certain intervals (for example, 51 cM in E x P and 11 cM in E x CL, for the distance between the 45S ribosomal repeat and TGlB on chromosome 2). In total, the markers scored in both crosses span 852 cM in the E x CL map versus 1103 cM in the E x P map, a highly significant ($P < 0.01$) difference. Skewed segregation ($P < 0.05$) was detected for 48 of the 70 markers, comprising 21 distinct regions distributed over all 12 chromosomes. The heterozygote (E/CL) was overabundant in 12 cases, whereas in nine cases the homozygote (E/E) was favoured. Overall, the effects of skewing approximately cancelled each other out: on average, the back-cross contained the expected 75% E genome (Figure 2).

d. Interval mapping of QTLs

Next the question of mapping the Mendelian factors that underly continuous variation in fruit mass, soluble-solids concentration and pH was addressed. The method of maximum likelihood and lod scores, commonly used in human linkage analysis has recently been adapted to allow interval mapping of QTLs. Ott, J., Analysis of Human Genetic Linkage (Johns Hopkins, Baltimore, 1985); Lander, E.S. and D. Botstein, Genetics, 121:185-199 (1989). Both were used here. At each position in the genome one computes the 'most likely' phenotypic effect of a putative QTL affecting a trait (the effect which

-14-

maximizes the likelihood of the observed data arising) and the odds ratio (the chance that the data would arise from a QTL with this effect divided by the chance that it would arise given no linked QTL). The lod score, defined as the \log_{10} of the odds ratio, summarizes the strength of evidence in favour of the existence of a QTL with this effect at this position; if the lod score exceeds a pre-determined threshold, the presence of a QTL is inferred. The traditional approach to mapping QTLs, as described by Tanksley et al. and Edwards et al., involves standard linear regression, which accurately measures the effect of QTLs falling at marker loci only, underestimating the effects of other loci in proportion to the amount of recombination between marker and QTL.

15 Tanksley, S.D. et al., Heredity, 49:11-25 (1982); Edwards, M.D. et al., Genetics, 116:113-125 (1987). In contrast, interval mapping allows inference about points throughout the entire genome and avoids confounding phenotypic effects with recombination, by using

20 information from flanking genetic markers. In the special case when a QTL falls exactly at a marker locus, interval mapping reduces to linear regression. A computer program, MAPMAKER-QTL, was written to implement interval mapping.

25 Due to the large number of markers tested, an extremely high lod score threshold must be adopted to avoid false positives. Given the genetic length of the tomato genome and the density of markers used, a threshold of 2.4 gives a probability of less than 5% that

30 even a single false positive will occur anywhere in the genome. Lander, E.S. and D. Botstein, Genetics, 121:185-199 (1989). This is approximately equivalent to

-15-

requiring the significance level for any single test to be 0.001.

QTL likelihood maps, showing how lod scores for fruit mass, soluble-solids concentration and pH change as one moves along the genome, reveal multiple QTLs for each trait and estimate their location to within 20-30 cM. (Figure 3).

Factors for fruit mass were found on six chromosomes (1, 4, 6, 7, 9 and 11). In each case, CL alleles decrease fruit mass (by 3.5 to 6.0 g), adding to a total reduction of 28.1 g inferred for back-cross progeny carrying a CL allele at all six loci. This accounts for about half of the approximately 60 g difference between E and CL.

Factors for soluble-solids concentration were found on four chromosomes (3, 4, 6 and 7). In each case, CL alleles elevate soluble-solids concentration (by 0.83 to 1.89 °Brix), adding to a total of 4.57 °Brix (versus a difference of approximately 5° Brix between the parental strains). This large effect in the backcross is consistent with previous reports that high soluble-solids concentration exhibits dominance and overdominance. Rick, C.M., Hilgardia, 42:493-510 (1974); Tanksley, S.D. and J. Hewitt, J. Theor. Appl. Genet., 75:811-823 (1988). The QTL alleles for both fruit mass and soluble-solids concentration all produce effects in the direction predicted by the difference between the parental strains.

Factors for pH were found on five chromosomes (3, 6, 7, 8 and 10). In addition, the lod score for a putative QTL on chromosome 9 fell just below the threshold. Because the parental strains do not differ greatly in pH, it was suspected that CL alleles might not all produce effects in the same direction. In fact, pH

-16-

was increased by four QTLs and decreased by two, including the likely QTL on chromosome 9. This provides a genetic explanation for the observation that many back-cross progeny exhibited more extreme phenotypes than the parental strains (Figure 1), a phenomenon known as transgression. Simmonds, N.W. Principles of Crop Improvement, 82-85 (Longman, NY, 1981).

Together, the QTLs identified for fruit mass, soluble-solids and pH account for 58%, 44% and 48%, respectively, of the phenotypic variance among the back-cross progeny, with another 13%, 9% and 11% attributable to environment.

The numbers of QTLs reported for each trait must be considered a minimum estimate. Because an extremely stringent threshold was used to avoid any false positives, some sub-threshold effects probably represent real QTLs. For example, the regions near TG19 on chromosome 1, CD41 on chromosome 5 and TG68 on chromosome 12 may affect soluble-solids concentration and merit further attention in larger populations. Similarly, the region near the μ locus on chromosome 10 may contain an additional QTL affecting pH (See Example 1). Moreover, one cannot rule out the presence of many additional QTLs with tiny phenotypic effects, which has been postulated in evolutionary theory and supported by some experimental evidence. Lande, R., Heredity, 50:47-65 (1983); Shrimpton, A.E. and A. Robertson, Genetics, 11:445-459 (1988). Also, it is conceivable that some of the apparent QTLs actually represent several closely-linked QTLs, each with small phenotypic effects in the same direction. Such a phenomenon might arise particularly in regions of genetic map compression. Finally, the QTL mapping here applies strictly only to the specific

-17-

environment tested and to heterozygosity for CL alleles. In principle, homozygosity for CL alleles could have been studied by using an F₂ self between E and CL, but in practice too many of the progeny are sterile.

05 Some regions of the genome clearly exert effects on more than one trait (for example, chromosome 6; Figure 3), providing a genetic explanation for at least some of the correlation between the traits. Although the present data are insufficient to distinguish between pleiotropic effects of a single gene and independent effects of tightly-linked loci, the frequent coincidence of QTL locations for different traits makes it likely that at least some of the effects are due to pleiotropy.

15 The region near sp on chromosome 6 has the largest effects on soluble solids and pH, as well as a substantial effect on fruit mass. The sp gene affects plant-growth habit: the dominant CL allele causes continuous apical growth (indeterminate habit), whereas the recessive E allele causes termination in an inflorescence ('determinate' or 'self-pruning' habit). Yeager, A.F., J. Hered., 18:263-265 (1927). Although indeterminacy has been reported previously by Emery et al. to elevate both fruit mass and soluble-solids concentration within L. esculentum, it is associated here with reduced fruit mass in both E x CL and another interspecific cross (E x L. cheesmanii). Emery, G.C. and H.M. Munger, J. Am. Soc. Hort. Sci., 95:410-412 (1966). These differing results might be due to a second, tightly-linked locus or to unlinked modifier genes.

30 Overall, pairwise epistatic interactions between intervals were not common (about 5% of two-way analysis-of-variance tests were significant at 0.05).

-18-

An interesting exception was the region near TG16 on chromosome 8, at which the CL allele significantly enhanced the effect of three of the four QTLs for soluble-solids concentration. TG16 also showed the most
05 extreme segregation distortion of any marker scored (about 4:1 in favour of the E/E homozygote) and is in a region known to exhibit skewed segregation in back-crosses to other green-fruited tomato species. Zamir, D. and Y. Tadmor, Bot. Gaz., 147:355-358 (1986);
10 Tanksley, S.D. In: Isozymes in Plant Genetics and Breeding, (eds. Tanksley, S.D. and T.J. Orton) 331-338 (Elsevier, Amsterdam, 1983). The unusual properties of this region of CL clearly merit further study.

The QTLs identified here may well differ from those
15 that would be fixed by repeated back-crossing with continuing selection for a trait, a classical method for introgressing quantitative traits. Work on LA1563, a strain with increased soluble solids produced through back-crossing a different strain of E to CL has provided
20 some suggestive evidence. Rick, C.M., Hilgardia, 42:493-510 (1974). By surveying RFLPs, Tanksley and Hewitt recently found that LA1563 has maintained three separate regions from CL: near CD56 on chromosome 10, near Got2 on chromosome 7 and near TG13 on chromosome 7.
25 Here, above-threshold effects were detected in the last of these three regions only (which, interestingly, failed to show effects on soluble solids in a single-environment test by Tanksley and Hewitt). Tanksley, S.D. and J. Hewitt, J. Theor. Appl. Genet., 75:811-823 (1988).
30 Moreover, QTLs affecting soluble-solids concentration were detected in regions that did not seem to be retained. Unfortunately, the results of the two experiments are not directly comparable due to the use of

-19-

a different strain by Rick, possible environmental differences between the experiments, the possibility that small CL fragments containing QTLs went undetected in LA1563, the possibility that the region near TG13 retained in LA1563 may not contain the QTL detected here and the possibility that some of the sub-threshold effects are real. Although more detailed studies are clearly needed, it is interesting to speculate about why repeated back-crossing may fix a narrower class of QTLs than found by QTL mapping. Because such breeding programs demand horticultural acceptability, they are likely to select against otherwise-desirable QTLs which are closely linked to undesirable effects from the wild parent. If such QTLs can first be identified by mapping it may be feasible to remove linked deleterious effects by recombination.

Once several QTLs with relatively large effects have been mapped, crosses can be used to isolate them in near-isogenic lines. These lines can be used to characterize the QTLs in various dosages, genetic backgrounds, environments and combinations. By re-assembling selected CL alleles in an otherwise E genotype, it should be possible to engineer an agriculturally-useful tomato with a higher yield of soluble solids.

e. Overview of Key Considerations

Although it has long been recognized that quantitative traits often arise from the combined action of multiple Mendelian factors, only recently has it become practical to undertake systematic mapping of such QTLs. While such investigations will by no means be easy, the methodology developed here should increase

-20-

their accuracy and efficiency. Specifically, by integrating information from genetic markers spaced throughout a genome, the method of interval mapping described above allows (i) efficient detection of QTLs
05 while limiting the overall occurrence of false positives; (ii) accurate estimation of phenotypic effects of QTLs; and (iii) localization of QTLs to specific regions (Figure 6). Beyond the increased efficiency due to interval mapping, the strategy of selective genotyping
10 (when applicable) can further reduce the number of progeny that must be genotyped in order to detect a QTL. Together, the methods lead to a reduction of up to 7-fold in the number of progeny to be genotyped. Finally, additional savings may be achieved via progeny testing
15 and simultaneous search. The main considerations in designing a cross for genetic dissection of a quantitative trait are as follows:

1. Designing a cross: Strains can be chosen to maximize the chance that they segregate for QTLs having
20 relatively large phenotypic effects, thereby allowing mapping with a manageable number of progeny. The ideal situation occurs when (a) the phenotypic difference D between the strains is large compared to the environmental or within-strain standard deviation σ_E ; (b)
25 breeding experiments indicate that the number k of effective factors given by Wright's formula is small; and (c) the strains are the result of selective breeding for the trait.

2. Specifying the minimum phenotypic effect the cross will be designed to detect: Once the strains have been
30 chosen, the experimenter must specify the minimum phenotypic effect δ that the cross will be designed to detect. When using strains resulting from selection, a

-21-

choice of δ in the range of between $\frac{1}{2}(D/k)$ and (D/k) should ensure that QTLs accounting for much of the phenotypic difference will be detected. When using arbitrary strains, the same choice of δ can be used, although the presence of QTLs with this effect is not guaranteed.

3. Calculating the number of backcross progeny to be phenotyped: The number N of backcross progeny that should be genotyped can then be calculated based on the spacing d between genetic markers in the map, the appropriate threshold T for the LOD score, and the desired probability β of success, assuming either (i) the traditional method of analysis involving single markers and genotyping all progeny or (ii) interval mapping and selective genotyping of the 5% most extreme progeny. Figure 10 shows N as a function of the fraction of variation v explained by the QTL (where $v = \delta^2/16\sigma_{B1}^2$), while Figure 11ab shows N as a function of the phenotypic difference D between the strains and the number k of effective factors. Together, interval mapping and selective genotyping reduce the number of progeny to be genotyped by up to 7-fold. Both Figures 10 and 11 assume that $d = 20$ cM, $T = 2.5$, and $\beta = 0.50$, and Figure 11 assumes that the QTLs have equal phenotypic effects. For instances in which different assumptions are made, the following modifications are made: multiply by 4 to allow for QTLs having half the average effect; multiply by approximately $(1.25)(1-2\theta)^2/(1-\psi)$ to allow for markers every d cM (θ and ψ are the recombination fractions corresponding to $\frac{1}{2}d$ and d cM, respectively; multiply by approximately 1.50 to allow for a 90% chance of success; multiply by approximately 1.50 to allow for a 90% chance of success; multiply by $T/2.5$ to allow for a low LOD

-22-

threshold of T; and multiply by about 55% if an F2 intercross is used instead of a backcross. As a rule of thumb, it appears practical to map QTLs when the phenotypic difference D measured in environmental standard deviations is on the order of the number k of effective factors segregating.

The Spontaneous Hypertensive rat (SHR) strain (Tanase, H. et al., Japan. Circulation Journal, 34:1197-1212 (1970)), was derived from the Wistar-Kyoto rat (WKY) strain by selective breeding for high systolic blood pressure followed by inbreeding. Blood pressure in SHR is about 3 standard deviations higher than in WKY, while the number of effective factors was estimated at k about 3. Assuming that the rat genome is about 1500 cM and that a 20 cM map RFLP is available, the appropriate LOD threshold would be about 2.7 (see Figure 7). Using the traditional approach, one would need about 325 backcross progeny or about 175 F2 intercross progeny. With interval mapping, these become about 275 and 145. If it were practical to grow a larger population but genotype only those progeny with the 5% most extreme blood pressures, the number of progeny to genotype could be reduced to about 55 and 30, respectively.

In addition to SHR, a number of other genetically hypertensive strains of rat and mouse have been desired, with estimated number of factors between 2 and 5 (DeJong 1984). Comparison of these strains would elucidate the number and location of the most important genes controlling naturally-occurring variation for blood pressure--at least in rodent population. Such information might shed light on hypertension in humans as well.

- 23 -

The availability of complete RFLP linkage maps makes it possible to dissect quantitative traits into discrete genetic factors, thereby unifying two historically-separated areas of genetics. Once QTLs have
05 been mapped, isogenic lines can be rapidly constructed differing only in the region of the QTL by using the RFLPs to select for the desired region and against the remainder of the genome. Soller, M. & J.S. Beckmann, Theor. Appl. Genet., 47:179-190 (1983); Paterson, A.H. et al., Submitted, 1988.
10

f. Application of QTL Mapping

The general approach of QTL mapping is broadly applicable to a wide range of biological endeavors. For example, in agriculture, it might be desirable to
15 transfer to domestic strains many quantitative traits harbored in wild species, including resistance to diseases and pests, tolerance to drought, heat, cold and other adverse conditions, efficient use of resources and high nutritional quality. Rick, C.M., In: Genes, Enzymes, and Populations (Ed. A.M. Srb) 255-268 (Plenum, NY, 1973); Harlan, J.R., Crop Sci., 16:329-333 (1976).
20 In mammalian physiology, selective breeding has generated rodent strains which differ greatly in quantitative traits, such as hypertension, atherosclerosis, diabetes, predispositions to cancer, drug sensitivities and various
25 behavioural patterns. Information on the number, location and nature of these QTLs would be of value in medicine. Festing, M.F.W., Inbred Strains in Biomedical Research, (Oxford, 1979). In evolutionary biology, the
30 process of speciation can be investigated by studying the number and nature of genes underlying reproductive

-24-

isolation. Coyne, J.A. and B. Charlesworth, Heredity,
57:243-246 (1986).

The availability of detailed RFLP linkage maps makes
it possible to dissect quantitative traits into discrete
05 genetic factors (QTLs): all regions of a genome can be
assayed and accurate estimates of phenotypic effects and
genetic position derived from interval analysis.
Tanksley, S.D. et al., R. Proc. 18th Stadler Genet.
Symp., In Press; Helentjaris, T., Trends in Gen.,
10 3(8):217-221 (1987); Landry, B.S. et al., Theor. Appl.
Genet., 74:646-653 (1987); Burr, B. et al., Genetics,
118:519-526 (1988); Chang, C. et al., Proceedings of the
National Academy of Sciences, U.S.A., In Press; McCouch,
S.R. et al., Theor. Appl. Genet., In Press; Kosambi,
15 D.D., Ann. Eugen., 12:172-175 (1944). Once QTLs are
mapped, RFLP markers permit genetic manipulations such as
rapid construction of near-isogenic lines: flanking
markers may be used to retain the QTL and the study of
the remaining markers may be used to speed progress by
20 identifying individuals with a fortuitously high
proportion of the desired genetic background (see Example
1). Using isogenic lines, the fundamental tools of
genetics and molecular biology may be brought to bear on
the study of QTLs, including testing of complementation,
25 dominance and epistasis; characterization of
physiological and biochemical differences between
isogenic lines; isolation of additional alleles by
mutagenesis (at least in favourable systems); and
physical mapping and molecular cloning of genetic factors
30 underlying quantitative traits.

-25-

EXAMPLE 1 Interspecies Back-cross and Assessment of Progeny

The tomatoes were grown in the field at Davis, California, in a completely randomized design including 05 237 BC plants (with E as the recurrent pistillate parent), as well as E, CL and the F_1 are controls. Neither CL nor the F_1 progeny matured completely, as is typical in the central valley of California. Among the BC plants, six failed to mature and 12 produced too few 10 fruit to assay reliably for quantitative traits. The absence of quantitative trait data for these few progeny should yield at most a slight bias in the analyses. The frequency distribution for fruit masses, soluble solids concentration ($^{\circ}$ Brix, a standard refractometric measure 15 primarily detecting reducing sugars, but also affected by other soluble constituents; 1° Brix is approximately 1% w/w) and pH in the E parental strain and in the backcross (BC) progeny are shown in Figure 1. Means and standard deviations for the distributions of the E parental strain 20 (E filled bars) and the BC progeny (BC open bars) appear in the upper right of each histogram. The distributions for soluble-solids concentration and pH are approximately normal. The distribution of the BC progeny for fruit weight is clearly skewed; \log_{10} (fruit mass) was studied 25 throughout to achieve approximate normality ($E = 1.81 \pm 0.07$; $BC = 1.20 \pm 0.19$). Wright, S., Evolution and the Genetics of Populations, (Univ. of Chicago, Chicago, 1968). The proportion of variance due to environment was estimated to be the square of the ratio of the standard 30 deviations (E/BC), for log-mass, solids and pH. Figure 2 shows the distribution of percentage of recurrent parent (E) genotype in the 237 back-cross progeny, estimated on the basis of the marker genotypes and their relative distances. Determination of marker genotypes was as

-26-

previously described. Tanksley, S.D. and J. Hewitt, Theor. Appl. Genet., 75:811-823 (1988). Estimates of the percentage of recurrent parent genome were produced by the recently-developed computer program HyperGene™.

05 Although the average agreed closely with the Mendelian expectation of 75% for a back-cross, values for individual plants ranged from 59% to over 90%. The distribution of the proportion of recurrent-parent genome agrees with the mathematical expectation. Franklin,

10 L.A., Theor. Populat. Biol., 11:60-80 (1977); Stam, P., Genet. Res., 25:131-155 (1980). The individual with >90% E appears to carry only five fragments from CL (ranging from 9 to 47 map units in length) and could be returned to essentially 100% E with two additional back-crosses of

15 about 550 plants. This is far more rapid than the 6-8 back-crosses routinely used to eliminate donor genome in the absence of markers.

QTL likelihood maps indicating lod scores for fruit mass (solid lines and bars), soluble-solids concentration

20 (dotted lines and bars) and pH (hatched lines and bars), throughout the 862 cM spanned by the 70 genetic markers are shown in Figure 3. The RFLP linkage map used in the analysis is presented along the abscissa, in Kosambi cM. Kosambi, D.D., Ann. Eugen., 12:172-175 (1944). The order

25 of the markers agrees with the previously-published map of the E x P cross, except for three inversions of adjacent markers: (TG24-CD15), (TG63-CD32B) and (TG30-TG36). Tanksley, S.D. et al., R. Proc. 18th Stadler Genet. Symp., In Press. In the first case,

30 re-analysis of the E x P data with MAPMAKER indicates that the order shown here is the more likely order, in both E x P and E x CL. Lander, E.S. et al., Genomics, 1:174-181 (1987); Proc. Natl. Acad. Sci., USA, 84:2363-2367 (1987). For the other two, the orders shown

-27-

here are more likely in E x CL by odds of $10^4:1$ and $10^7:1$, but the inverse is more likely in E x P by 11:1 and 8:1 odds. These differences will be investigated in a larger E x P population. Soluble-solids concentration and pH were analyzed in °Brix and pH units, respectively; allele effects on fruit mass are presented in g; log transformation of fruit mass was used in all analyses to achieve approximate normality. The maximum likelihood effect of a putative QTL, as well as the lod score in favour of the existence of such a QTL, have been determined at points spaced every 1 cM throughout the genome, according to the method described herein and a smooth curve plotted through the points. The height of the curve indicates the strength of the evidence (\log_{10} of the odds ratio) for the presence of a QTL at each location and not the magnitude of the inferred allelic effect. The horizontal line at a height of 2.4 indicates the stringent threshold that the lod score must cross to allow the presence of a QTL to be inferred, as described herein.

Information about the likely position of the QTL can also be inferred from the curve. The maximum likelihood position of the QTL is the highest point on the curve. Bars below each graph indicate a 10:1 likelihood support interval for the position of the QTL (the range outside which the likelihood falls by a lod score of 1.0), whereas the lines extending out from the bars indicate a 100:1 support interval. Phenotypic effects indicated beside the bars are the inferred effect of substituting a single CL allele for one of the two E alleles at the QTL.

Several regions show sub-threshold effects on one or more traits (chromosome one near TG19, chromosome five near TG34 and chromosome 12 near TG68) which may

-28-

represent QTLs; this requires additional testing. The region near TG68 may be particularly interesting, as it is the only instance found where the CL allele seems to decrease soluble-solids concentration (by about
 05 0.7°Brix). In the case of chromosome 10, the lod score for pH crosses the significance threshold in two places. Controlling for the presence of a QTL near CD34A, testing for the presence of a second QTL near μ was carried out, by comparing the maximum lod scores assuming the presence
 10 of only the first QTL to the maximum lod score assuming the presence of two QTLs). Allowing for a QTL in the region of CD34A, the residual lod score near μ falls below the required threshold. Thus, the evidence is not yet sufficient to support the presence of a QTL near μ .

15 Methods. The lod score and the maximum likelihood estimate (MLE) of the phenotypic effect at any point in the genome is computed assuming that the distribution of phenotypes in the BC progeny represents a mixture of two normal distributions (of equal variance) with means
 20 depending on the genotype at a putative QTL at the given position. (Note that QTLs are considered individually and there is no assumption that different QTL effects can be added, except in studying the possibility of two QTLs on chromosome 10 affecting pH. Specifically, at a given
 25 position in the genome, the likelihood function for individual i with quantitative phenotype ϕ is given by

$$L_1(\alpha, \sigma) = (2\pi\sigma^2)^{-1/2} (p_1 \exp(-\phi^2 / 2\sigma^2) + p_2 \exp(-(\phi_1 - \alpha)^2 / 2\sigma^2))$$
 where α is the effect of substituting a CL allele for an E allele at a putative QTL in the given
 30 position, σ^2 is the phenotypic variance not attributable to the QTL and p_1 and p_2 are the probabilities that individual i has genotype E/E and E/CL, respectively, at

-29-

the QTL (which can be computed on the basis of the genotypes at the flanking markers and the distance to the flanking markers). The likelihood function for the entire population is $L = 11 L_1$. Also α^* and σ^* denote the MLEs allowing the possibility of a QTL at the location (the values which maximize L) and σ^{**} denotes the MLE of σ , subject to the constraint that no QTL is linked ($\alpha = 0$). The lod score is then given by $\log_{10} L(\alpha^*, \sigma^*) / L(0, \sigma^{**})$. This method for QTL mapping is developed more fully in Example 2.

EXAMPLE 2 Analytical Methods Used in Mapping Quantitative Traits Using RFLP Linkage Maps

The following is a description of a set of analytical methods that modify and extend the classical theory for mapping discrete Mendelian factors underlying quantitative traits, referred to as quantitative trait loci (QTLs). These include: (i) a method of identifying promising crosses for QTL mapping by exploiting a classical formula of Sewall Wright; (ii) a method (interval mapping) for exploiting the full power of RFLP linkage maps by adapting the approach of LOD score analysis used in human genetics, to obtain accurate estimates of the genetic location and phenotypic effects of QTLs; and (iii) a method (selective genotyping) that allows a reduction of up to 7-fold in the number of progeny that need to be scored with the DNA markers. Figures 10 and 11 are graphs that allow geneticists to estimate, in any particular case, the number of progeny required to map QTLs underlying a quantitative trait.

(i) Identification of promising crosses for QTL mapping. Genetic dissection of a quantitative trait will

-30-

succeed only when some of the QTLs segregating in the cross have relatively large phenotypic effects. It has been shown, through use of a classical formula of Sewall Wright, that it is often possible to recognize such
05 crosses in advance and thereby to ensure that QTLs will in fact be identified.

The basic methodology for mapping QTLs involves arranging a cross between two inbred strains differing substantially in a quantitative trait: segregating
10 progeny are scored both for the trait and for a number of genetic markers. Typically, the segregating progeny are produced by a B1 backcross (F1 x Parent) or an F2 intercross (F1 x F1). For simplicity, only the backcross will be discussed in detail. As noted below, the F2
15 intercross is completely analogous and requires only about half as many progeny.

Definitions and assumptions: Let A and B be inbred strains differing for a quantitative trait of interest, and suppose that a B1 backcross is performed with A as
20 the recurrent parent. Let

$$(\mu_A, \sigma_A^2), (\mu_B, \sigma_B^2), (\mu_{F1}, \sigma_{F1}^2) \text{ and } (\mu_{B1}, \sigma_{B1}^2)$$

denote the mean and variance of the phenotype in the A, B, F1 and B1 populations, respectively (see Figure 4). Let $D = \mu_B - \mu_A > 0$ denote the phenotypic difference
25 between the strains. The cross will be analyzed under the classical assumption that the phenotype results from summing the effects of individual QTL alleles, and then adding normally-distributed environmental (i.e., non-genetic) noise. (Mather, K. and J.L. Jinks,
30 Biometrical Genetics, Cornell University Press, Ithaca, NY (1971); Falconer, D.S., Introduction to Quantitative

-31-

Genetics, Longman, London (1981)). In particular, complete codominance and no epistasis are assumed. These assumptions imply that:

$$\begin{aligned} \mu_{F1} &= \frac{1}{2}(\mu_A + \mu_B), & (1a) \\ \mu_{BC1} &= \frac{1}{2}(\mu_A + \mu_{F1}), \text{ and} & (1b) \\ \sigma_A^2 = \sigma_B^2 = \sigma_{F1}^2 &< \sigma_{BC1}^2, & (1c) \end{aligned}$$

The variances within the A, B and F1 populations equal the environmental variance, σ_E^2 , among genetically identical individuals, while the variance within the B1 progeny also includes genetic variance $\sigma_G^2 = \sigma_{B1}^2 - \sigma_E^2$. Frequently, phenotypic measurements must be mathematically transformed so that parental phenotypes are approximately normally distributed and the relations (1abc) are approximately satisfied. For example, Wright (1968) obtained an excellent fit to the theory by applying a log-transformation (appropriate when variances scale with the mean) to tomato fruit weight.

By the phenotypic effect δ of a QTL, is meant the additive effect of substituting both A alleles by B alleles. A single allele has effect $\frac{1}{2}\delta$, since codominance is assumed. In a backcross, the segregation of a QTL with effect δ contributes an amount $\delta^2/16$ to the genetic variance σ_G^2 . The variance explained by the QTL is written $\sigma_{\text{exp}}^2 = \delta^2/16$, while the residual variance is $\sigma_{\text{res}}^2 = \sigma_{B1}^2 - \sigma_{\text{exp}}^2$.

Choosing Strains

The ability to map QTLs underlying a quantitative trait depends on the magnitude of their phenotypic effect: the smaller the effect that one wishes to detect, the more progeny will be required. Before

-32-

attempting genetic dissection of a quantitative trait, it would thus be desirable to identify crosses segregating for QTLs with relatively large phenotypic effects and to estimate the magnitude of the effects. In fact, this can
05 often be accomplished by exploiting a classical formula of Wright.

Wright (quoted by Castel 1921; Wright 1968) proved that the number of QTLs segregating in a backcross between two strains with phenotypic difference D can be
10 estimated by the formula:

$$k = D^2 / 16\sigma_G^2 \quad (2)$$

provided that the following assumptions hold: (i) the QTLs have effects of equal magnitude; (ii) the QTLs are
unlinked; and (iii) the alleles in the high strain all
15 increase the phenotype, while those in the low strain decrease the phenotype. The estimate k is called the number of effective factors in the cross. If the assumptions are satisfied, then each QTL affects the phenotype by (D/k) and explains $(1/k)$ of the genetic
20 variance in the backcross.

Unfortunately, if these assumptions are not satisfied (as will be likely in practice), the number of effective factors k may seriously underestimate the number of QTLs. In principle, the number of QTLs is
25 unlimited. In this case, must there exist any QTLs affecting the phenotype by (D/k) ? More generally, for any $0 \leq \epsilon \leq 1$, must there exist QTLs affecting the phenotype by $\epsilon(D/k)$? And, how must of the total phenotypic difference D and the genetic variance σ_G^2 can
30 be attributed such QTLs? Proposition 1 (proven in Appendix [A1]) supplies an answer.

- 33 -

Proposition 1.

Consider a cross in which the phenotype difference between the strains is D and the number of effective factors is k . Assume that the QTLs are unlinked and that the alleles in the 'high' strain all increase the phenotype. No matter how many QTLs are segregating and no matter what their individual phenotype effects, the sets of QTLs that alter the phenotype by at least $\epsilon(D/k)$ must together account for a fraction $\geq D_\epsilon$ of the total phenotypic difference D between the strains and must together explain a fraction $\geq V_\epsilon$ of the genetic variance in the second generation, where

$$D_\epsilon = [k\epsilon + (1-\epsilon)k + k\epsilon^2]/k \text{ and } V_\epsilon = 1 - \epsilon(1-D_\epsilon).$$

Considering the case $\epsilon=1$, the proposition states that the QTLs with phenotypic effect (D/k) must account for a phenotypic difference of at least (D/k) . In other words, there must exist at least one QTL having phenotypic effect $\geq (D/k)$.

Consider the search for QTLs with somewhat smaller effects. How much of the phenotypic difference can be attributed to QTLs with effect $\geq \frac{1}{2}(D/k)$? Taking $\epsilon = \frac{1}{2}$ and considering various values of k , results in the following:

-34-

		Minimum proportion of phenotypic difference D accounted for by QTLs with effect $\geq \frac{1}{k}(D/k)$	Minimum proportion of genetic variance σ^2_G explained by QTLs with effect $\geq \frac{1}{k}(D/k)$
05	2	64%	82%
	3	50%	75%
	4	42%	71%
	5	37%	69%

A small value of k thus implies that the cross must be
 10 segregating for QTLs with relatively large effects
 ($\geq \frac{1}{k}(D/k)$), which together account for a substantial
 proportion of the phenotypic difference and explain a
 substantial proportion of the genetic variance in the
 backcross.

15 In other words, Wright's formula can be used to
 indicate the presence of some QTLs with large effects--
 even though the number k of effective factors may not be
 a reliable estimate of the total number of QTLs. Note
 that Proposition 1 provides only worst-case lower bounds:
 20 in general, the QTLs with large effects will have an even
 greater effect.

How serious a limitation is posed by the two
 assumptions remaining in Proposition 1?

(i) The first assumption is not essential:
 25 admitting the possibility of linked QTLs simply allows
 that some large QTL effects may eventually prove to be
 due to several nearby genes. Such questions may be
 safely neglected at first.

(ii) The second assumption is more important.
 30 Fortunately, it is possible to choose crosses in which it
 is likely to be satisfied. The ideal situation would be

-35-

two strains arising from brief, intense artificial selection for and against the trait in an outbred population, followed by inbreeding: in such a case, classical selection theory shows that a "high" strain is unlikely to fix a "low" allele at QTLs with relatively large effect; moreover, the force of selection will be greatest on the QTLs with the largest effects. (Falconer, D.S., Introduction to Quantitative Genetics, Longman, London (1981)). Many such strains have been developed to study various physiological traits. As a reasonable alternative, one could use strains that appear to have resulted from natural selection for the trait.

Judicious choice of strains can essentially assure that some QTLs will be detected in a reasonable progeny size that can be calculated in advanced. When studying strains resulting from selection, a reasonable approach to mapping QTLs would be to use enough progeny to map QTLs having effect δ between $\frac{1}{2}(D/k)$ and (D/k) . Of course, one could choose to study more progeny and might well be rewarded with the detection of QTLs with smaller effects.

Unselected strains exhibiting extreme phenotypic difference may also merit attention: QTLs with large effects may well be segregating, despite the lack of a mathematical guarantee. When there is clear evidence of both high and low alleles within a strain--as when many segregating progeny exhibit phenotypes more extreme than either parent--the analysis above does not apply; the detection level for QTLs must be chosen somewhat arbitrarily. When there is no such evidence, one might proceed as above in choosing a progeny size. Although the existence of QTLs having a given effect is no longer

-36-

assured, the methods described below for detecting QTLs in a cross are unchanged.

Assuming that the desired detection level δ has been chosen, next considered are the method for mapping QTLs
05 and the number of progeny required.

(ii) Exploit the full power of complete linkage maps. The traditional approach to mapping QTLs involves studying single genetic markers one-at-a-time. Sax, K., Genetics, 8:552-560 (1923; Soller, M. and T. Brody, 10 Theor. Appl. Genet., 47:35-39 (1976). In general, the drawbacks of the method include that (a) the phenotypic effects of QTLs are systematically underestimated, (b) the genetic locations of QTLs are not well resolved because distant linkage cannot be distinguished from 15 small phenotypic effect, and (c) the number of progeny required for detecting QTLs is larger than necessary. Adapting the method of LOD scores used in human genetic linkage analysis, it has been possible to remedy these problems by the approach of interval mapping of QTLs. In 20 addition, the traditional approach neglects the problem that testing many genetic markers increases the risk that false positives will occur. As described below, the appropriate degree of statistical stringency to prevent such errors in mapping QTLs has been determined.

25 Traditional Approach to Mapping QTLs

The traditional approach for detecting a QTL near a genetic marker involves comparing the phenotypic means for two classes of progeny: those with genotype marker AB, and those with marker genotype AA. The difference 30 between the means provides an estimate of the phenotypic effect of substituting a B allele for an A allele at the

-37-

QTL. To test whether the inferred phenotypic effect is significantly different from 0, one applies a simple statistical test--amounting to linear regression or one-way analysis of variance, under the assumption of normally-distributed environmental variance.

Consider a QTL that contributes σ_{exp}^2 to the genetic variance. Supposing that such a QTL were located exactly at a marker locus, the number of progeny required for detection would be approximately

$$(Z_{\alpha})^2 (\sigma_{\text{res}}^2 / \sigma_{\text{exp}}^2), \quad (3)$$

where this progeny size affords a 50% probability of detection if such a QTL is actually present and a probability α of a false positive if no QTL is linked. Here, Z_{α} is defined by the equation Probability ($z > Z_{\alpha}$) = α where z is a standard normal variable (i.e., Z_{α} is the number of standard deviations beyond which the normal curve contains probability α). Soller and Brody suggest allowing a false positive rate of $\alpha = 0.05$. Soller, M. and T. Brody, Theor. Appl. Genet., 47:35-39 (1976). For a given false positive rate, the required progeny size thus essentially scales inversely with the square of the phenotypic effect of the QTL or, equivalently, inversely with the variance explained.

Although it captures the key features of QTL mapping, the traditional approach has a number of shortcomings:

- (i) If the QTL does not lie at the marker locus, its phenotypic effect may be seriously underestimated. If the recombination fraction is θ , the inferred phenotypic effect of the QTL is biased downward by a factor of $(1-2\theta)$.

-38-

(ii) If the QTL does not lie at the marker locus, substantially more progeny may be required. In particular, the variance explained by the marker decreases by a factor of $(1-2\theta)^2$ and the number of
05 progeny consequently increases by a factor of $1/(1-2\theta)^2$. For an RFLP map with markers every 10, 20, 30 or 40 cM throughout the genome, the progeny size would need to be increased by 22%, 49%, 82% or 123%, respectively, to account for the possibility that the QTL might lie in the
10 middle of an interval (i.e., at the maximum distance from the nearest RFLP).

(iii) The approach does not define the likely position of the QTL. In particular, it cannot distinguish between tight linkage to a QTL with small
15 effect and loose linkage to a QTL with large effect.

(iv) The suggested false positive rate of $\alpha = 0.05$ neglects the fact that many markers are being tested. While the chance of a false positive at any given marker is only 5% the chance that a false positive will occur
20 somewhere in the genome is much higher.

These difficulties stem from the fact that single markers are analyzed one-at-a-time. To remedy these problems, the approach is generalized, as described in the following section, to make it possible to exploit the
25 full power of an RFLP linkage map to scan the intervals between markers as well.

QTL Mapping: Interval Mapping using LOD Scores

Method of maximum likelihood: The traditional
approach, involving linear regression of phenotype on
30 genotype, is a special case of the method of maximum likelihood. Formally, the phenotype ϕ_i and genotype g_i are assumed to be related by the equation

-39-

$$\phi_i = \alpha + bg_i + \epsilon,$$

where g_i is encoded as a (0,1)-indicator variable, ϵ is a random normal variable with mean 0 and variance σ^2 , and a , b , and σ^2 are unknown parameters. Here, b denotes the
 05 estimated phenotypic effect of a putative QTL.

The linear regression solutions (a^* , b^* , σ^{2*}) are, in fact, maximum likelihood estimates (MLEs) for the parameters--that is, they are the values which maximize the probability $L(a, b, \sigma^2)$ that the observed data would
 10 have occurred. Here,

$$L(a, b, \sigma^2) = \prod_i z((\phi_i - (a + bg_i)), \sigma^2), \quad (4)$$

where $z(x, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp(-x^2/2\sigma^2)$ is the probability density for the normal distribution with mean 0 and variance σ^2 . The MLEs are compared to the constrained
 15 MLEs obtained under the assumption that $b = 0$, corresponding to the assumption that no QTL is linked. These constrained MLEs are easily seen to be $(\mu_A, 0, \sigma_{BC1}^2)$. The evidence for a QTL is then summarized by the LOD score:

$$20 \quad \text{LOD} = \log_{10} (L(a^*, b^*, \sigma^{2*}) / L(\mu_A, 0, \sigma_{BC1}^2)),$$

essentially indicating how much more likely the data is to have arisen assuming the presence of a QTL than assuming its absence. (The choice of \log_{10} accords with longstanding practice in human genetics, although \log_e
 25 would be slightly more convenient below). Morton, N.E., Am. J. Hum. Genet., 7:277-318 (1955). If the LOD score exceeds a predetermined threshold T , a QTL is declared to be present. The important issues are: (1) What LOD

-40-

threshold T should be used, in order to maintain an acceptable low rate of false positives? (ii) What is the expected contribution to the LOD score (called the ELOD) from each additional progeny? The number of progeny
 05 required is then T/ELOD, to provide even odds of detecting the QTL with the desired false positive rate.

When only a single genetic marker is being tested, these equations are easily answered. (i) By a general result about maximum likelihood estimation in large
 10 samples, LOD is asymptotically distributed as $\frac{1}{2}(\log_{10} e) \chi^2$, where χ^2 denotes the chi-squared distribution with one degree of freedom. Kendall, M. and A. Stuart, The Advanced Theory of Statistics, Vol. 2, Griffin: London 1979). A false positive rate of α will
 15 thus result if the LOD threshold is chosen so that $T = \frac{1}{2}(\log_{10} e)(Z\alpha)^2$. For the 5% error rate suggested by Soller and Brody, the threshold is $T = 0.83$. The question of the appropriate threshold when many markers are being tested is postponed temporarily. (ii) For a
 20 QTL contributing σ_{exp}^2 to the backcross variance, the expected LOD score per progeny (ELOD) is

$$\text{ELOD} = \frac{1}{2} \log_{10}(1 + \sigma_{\text{exp}}^2 / \sigma_{\text{res}}^2) \quad (5a)$$

$$\approx \frac{1}{2}(\log_{10} e) (\sigma_{\text{exp}}^2 / \sigma_{\text{res}}^2) \quad (5b)$$

$$\approx 0.22 (\sigma_{\text{exp}}^2 / \sigma_{\text{res}}^2) \quad (5c)$$

25 where (5a) follows from well-known results about linear regression and (5b) follows from Taylor expansion for small values of $(\sigma_{\text{exp}}^2 / \sigma_{\text{res}}^2)$. Combining these two results, the number of progeny required so that the LOD score is expected to exceed T is

$$30 \quad T/\text{ELOD} = (Z_{\alpha})^2 (\sigma_{\text{exp}}^2 / \sigma_{\text{res}}^2) \quad (6)$$

-41-

This confirms that the maximum likelihood approach agrees with the result (3) from the traditional approach above, when examining effects at a single marker locus. The more general framework of maximum likelihood, however, allows
 05 the method to be generalized to several more complex situations described below.

(iii) Decrease the number of progeny to be genotyped. In typical cases, a reduction of up to 7-fold can be achieved by combining two approaches: interval
 10 mapping and selective genotyping. Selective genotyping involves growing a larger population, but genotyping only those individuals whose phenotypes deviate substantially from the mean. Additional methods for increasing the power of QTL mapping include reducing environmental noise
 15 by progeny testing and reducing genetic noise by studying several genetic regions simultaneously.

Interval mapping: If genetic markers have been scored throughout the genome, the method of maximum likelihood can be used as above to estimate the
 20 phenotypic effect and the LOD score for a putative QTL at any given genetic location. (Lander, E.S. and D. Botstein, Proceedings of the National Academy of Sciences, USA, 83:7353-7357 (1986); Lander, E.S. and D. Botstein, Cold Spring Harbor Symp. Quant. Biol., 51:49-62
 25 (1986)). The main difference is that the QTL genotype g_i for individual i is unknown: the appropriate likelihood function is therefore

$$L(a, b, \sigma^2) = \prod_i [G_i(0)L_i(0) + G_i(1)L_i(1)], \quad (7)$$

where $L_i(x) = z((\phi_i - (a+bx)), \sigma^2)$ denotes the likelihood
 30 function for the individual i assuming that $g_i = x$ and

-42-

$G_i(x)$ denotes the probability that $g_i = x$ conditional on the genotypes and positions of the flanking markers.

(Given a map function, G is easily computed. For example, if the flanking markers both have genotype AA in an individual and they lie at recombination fraction θ and θ' from the putative QTL, then the probability of the QTL genotype being AB is $\theta\theta'$, assuming no interference.) Note that (7) reduces to (4) in the special case that the QTL lies at a marker locus and the genotype g_i is thus known with certainty.

Finding the maximum likelihood solution (a^*, b^*, σ^{2*}) to (7) can be regarded as a linear regression problem with missing data: none of the independent variables (genotypes) are known; only probability distributions for each are available. Although standard computer programs for linear regressions cannot be used, techniques for maximum likelihood estimation with missing data have been developed in recent years. Little, R.J.A. and D.B. Rubin, Statistical Analysis with Missing Data, Wiley, NY (1987). By adapting the EM algorithm, a computer program MAPMAKER-QTL has been written to compute LOD scores for putative QTLs. Dempster, A.P. et al., J. Roy. Statist. Soc., 39:1-38 (1977); Lander, E.S. and P. Green, Proceedings of the National Academy of Sciences, USA, 84:2363-2367 (1987).

To illustrate the method, simulated data from many backcrosses has been analyzed. Figure 5 presents a QTL likelihood map, showing how the LOD score varies throughout a genome, for a simulated data set involving 250 backcross progeny segregating for five QTLs with various allelic effects. Based on the assumed genome size and density of markers, a LOD score of 2.4 is required (see below) for declaring the presence of a QTL.

-43-

In the example, the four largest QTLs are detected while the fifth does not attain statistical significance. The approximate position of the QTLs is indicated by one-lod confidence intervals, defined by the points on the genetic map at which the likelihood ratio has fallen by a factor of 10 from the maximum; such confidence intervals are frequently used in human genetics to indicate the probable position of genes. (Ott, J., Analysis of Human Genetic Linkage, Johns Hopkins University, Baltimore (1985)).

Among the advantages of the approach are:

- (i) The QTL likelihood map represents clearly the strength of the evidence for QTLs throughout the entire genome.
- (ii) In contrast to the traditional approach, the inferred phenotypic effects are asymptotically unbiased. This is an immediate consequence of the fact that they are MLEs for a correctly specified model. (Kendall, M. and A. Stuart, The Advanced Theory of Statistics, Vol. 2, Griffin: London (1979)).
- (iii) The probable position of the QTL is given by confidence intervals, indicating the range of points for which the likelihood ratio is within a factor of 10 (or 100, if desired) of the maximum.
- (iv) Interval mapping requires fewer progeny than the traditional approach for the detection of QTLs. In meioses in which the flanking markers do not recombine, the genotype of the QTL is known almost certainly--up to the chance of a double crossover (e.g., at most 1% in the case of a 20 cM RFLP map). In essence, the flanking markers can be thought of as a single tightly-linked virtual marker in such meioses. Supposing that genetic markers are available every d cM and considering the

-44-

(worst) case of a QTL in the middle of an interval, one can show (Appendix [A2]) that

$$ELOD_{\text{interval mapping}} = (1-2\theta)^2 ELOD / (1-\psi), \quad (8a)$$

where ψ is the recombination fraction corresponding to d cM, θ is the recombination fraction corresponding to $d/2$ cM, and ELOD is the expected LOD score for a marker located exactly at the QTL. By contrast, recall that

$$ELOD_{\text{single markers}} = (1-2\theta)^2 ELOD. \quad (8b)$$

Interval mapping thus decreases the required number of progeny by a factor of $(1-\psi)$ --which is exactly the proportion of meioses in which the flanking markers do not recombine. For maps with $d = 10, 20, 30$ and 40 cM, the savings are 9%, 16%, 23% and 28%, respectively.

(v) QTL likelihood maps can also be used to recognize a pair of linked QTLs, provided that they are not so close that recombination between them is very rare. Holding fixed the position of one QTL, the increase in LOD score caused by a second putative QTL can be computed for each position along the chromosome. An example is shown in Figure 6.

In addition to being tested on numerous simulated data sets, interval mapping has recently been applied to an interspecific backcross in tomato: six QTLs affecting tomato fruit weight, four QTLs affecting the concentration of soluble solids, and five QTLs affecting fruit pH were mapped to about 20-30 cM. (Paterson, A.H. et al., Submitted (1988)). In general, interval mapping should prove valuable for analyzing and presenting

-45-

evidence for QTLs, and for decreasing somewhat the number of progeny required to detect QTLs of a given magnitude.

Appropriate threshold for LOD scores: When an entire genome is tested for presence of QTLs, the usual
05 nominal significance level of 5% corresponding to a LOD score of 0.83 is clearly inadequate. Indeed, applying this standard would have resulted in a spurious QTLs being declared on chromosome 10 in Figure 5. The appropriate threshold depends on the size of the genome
10 and the density of markers genotyped.

To determine the correct LOD threshold, it is useful to consider two limiting situations: (i) the sparse-map case, in which consecutive markers are well-separated and (ii) the dense-map case, in which the spacing between
15 consecutive markers approaches zero. In each case, the issue is: If no QTLs are segregating, what is the chance that the LOD score will exceed the threshold T somewhere in the genome?

In the sparse-map case, occurrences of spuriously
20 high LOD scores are essentially independent. To achieve an overall significance level of α when M intervals are tested, a nominal significance level of α/M should be required for each individual test, corresponding to a LOD threshold of $4(\log_{10} e)(Z_{\alpha/M})^2$.

25 In the dense-map case, occurrences of spuriously high LOD scores at nearby markers are no longer independent events. Even if markers were typed continuously throughout the genome, there would be a maximum statistical penalty to be paid. In fact, as
30 shown in the Appendix [A3], in the limit of an infinitely dense-map and a large progeny size, the LOD score varies according to the square of an Orenstein-Uhlenbeck diffusion process. Well-known in physics and

-46-

engineering, the Ornstein-Uhlenbeck diffusion describes a particle executing Brownian motion while being coupled to the origin by a weak spring. The extreme value properties of this diffusion have been extensively
05 studied and the results immediately translate into statements about how high a LOD score will be expected by chance, given the size of the genome. (Leadbetter, M.R. et al., Extremes and related properties of random sequences and processes, Springer, NY (1983)).
10 Specifically, for a high threshold T , there is (see Appendix [A3]) the following result:

Proposition 2. Consider an organism with C chromosomes and genetic length G , measured in Morgans. When no QTLs are present, the probability that the LOD
15 score exceeds a high level T is $\approx (C + 2Gt) \chi^2(t)$, where $t = (4.6)T$. In order to make the probability less than α that a false positive occurs somewhere in the genome, the appropriate LOD threshold is thus $\approx T\alpha = 4.6t\alpha$, where t_α solves the equation $\alpha = (C + 2Gt_\alpha) \chi^2(t_\alpha)$.

20 For both the sparse-map and dense-map cases, a standard chi-square table may be used to calculate the LOD score threshold corresponding to a 5% chance that even a single false positive will occur. For intermediate situations, extensive numerical simulation
25 was used to determine the appropriate LOD thresholds as a function of genome size and marker spacing (Figure 7). Typically, a LOD score of between 2 and 3 is required to ensure an overall false positive rate of 5%. For instance, analyzing the domestic tomato ($C = 12$, $G = 11$)
30 with a 20 cM RFLP map requires a LOD threshold of 2.4-- equivalent to applying a nominal significance level of about $\alpha' = 0.001$ for each individual test performed. If the nominal 5% significance level ($\text{LOD} > 0.83$) were used

-47-

instead, the probability would exceed 90% that a false positive would arise somewhere in the genome. Indeed, a LOD score of 1.5 occurred by chance on chromosome 10 in the simulated data shown in Figure 5.

05 Number of progeny required: Given the ELOD for a QTL as a function of its phenotypic effect (Equation (8)) and the LOD threshold T (Figure 7) a progeny size of T/ELOD will ensure a 50% chance of detecting linkage to such a QTL no matter where it lies in the genome. If it
10 is desired to increase the chance of success to 100%, standard arguments show that the progeny size should be further increased by a factor of $[1 + (Z_{1-\beta}/Z_{\alpha'})]^2$, where α' is the nominal significance level corresponding to a LOD score of T. (Kendall, M. and A. Stuart, The Advanced
15 Theory of Statistics, Vol. 2, Griffin: London (1979)).

A technical note: The approximate progeny sizes given above (Equations 3, 5ab; 6, 8ab) are exact in the case of QTLs with small effects. Slight modifications are required for QTLs with large effects; see Appendix
20 [A4].

Increasing the Power of QTL Mapping

Although interval mapping increases the efficiency of QTL mapping somewhat, large numbers of progeny may still be required. Additional methods are available to
25 increase the power of QTL mapping, the most important of which is selective genotyping.

Selective genotyping of the extreme progeny: Some progeny contribute more linkage information than others. As a general principle, the individuals that provide the
30 most linkage information are those whose genotype can be most clearly inferred from their phenotype. For example, Lander and Botstein have pointed out that the vast

-48-

majority of linkage information about human diseases with incomplete penetrance comes from the affected individuals; since the genotype of unaffected individuals is uncertain, they provide relatively little information.

05 Lander, E.S. and D. Botstein, Cold Spring Harbor Symp. Quant. Biol., 51:49-62 (1986).

Applying this principle to quantitative genetics, the highest ELODs are provided by the progeny that deviate most from the phenotypic mean. When the cost of growing progeny is less than the cost of complete RFLP genotyping (as is frequently the case), it will thus be more efficient to increase the number of progeny grown but to genotype only those with the most extreme phenotypes. This increase in efficiency can be estimated as follows, with a more precise argument given in the Appendix [A5]. Since regression minimizes squared deviations from the mean, the ELOD conditional on an individual's phenotype ϕ is proportional to $(\phi - \mu_{B1})^2$. Thus, the proportion of individuals with extreme

15 as follows, with a more precise argument given in the Appendix [A5]. Since regression minimizes squared deviations from the mean, the ELOD conditional on an individual's phenotype ϕ is proportional to $(\phi - \mu_{B1})^2$. Thus, the proportion of individuals with extreme

20 phenotype ϕ such that $|\phi - \mu_{B1}| \geq L$ is

$$Q(L) = 2 \int_L^{\infty} z(x) dx,$$

while the proportion of the total linkage information contributed by such individuals is

$$\begin{aligned} S(L) &= 2 \int_L^{\infty} x^2 z(x) dx = Q(L) [1 + 2Lz(L)/Q(L)] \\ &= Q(L) [1 + L^2] \end{aligned} \quad (9)$$

25

using integration by parts and the approximation $z(L)/Q(L) = \frac{1}{2}L$ for large L . Accordingly, the same total linkage information would be obtained by growing a

-49-

population that was larger by a factor of $h(L) = 1/S(L)$, but only genotyping individuals with extreme phenotypes. The number of progeny to genotype would fall by a factor of $g(L) = S(L)/Q(L) = [1+L^2]$. Graphs of $Q(L)$, $S(L)$, $h(L)$ and $g(L)$ are shown in Figure 8. Results show that:

(i) Progeny with phenotypes more than 1 standard deviations from the mean comprise about 33% of the total population but contribute about 81% of the total linkage information. By growing a population that was only about 25% larger and genotyping only these extreme progeny, the same total linkage information would be obtained from genotyping only about 40% as many individuals.

(ii) Progeny with phenotypes more than 2 standard deviations from the mean comprise about 5% of the total population but contribute about 28% of the total linkage information. By growing a population that was about 3.6-fold larger and genotyping only these extreme progeny, the same total linkage information would be obtained from genotyping about 5.5-fold fewer individuals (since $h(2) = 3.6$ and $g(2) = 5.5$).

(iii) It is probably unwise to go beyond the 5% tails of distribution. From a practical point of view, true phenotypic outliers may represent artifacts. Moreover, the increase in population size required for $L > 2$ outweighs the decreased number of individuals to genotype.

The strategy of selective genotyping will substantially increase efficiency whenever growing and phenotyping additional progeny requires less effort than completely genotyping individuals at all RFLP markers, which is typically the case in many organisms.

It should be noted that standard computer programs for linear regression cannot be used (even for single

-50-

marker analysis) when only the extreme progeny have been genotyped; phenotypic effects would be grossly overestimated because of the biased selection of progeny. As in the case of interval mapping, missing-data methods are required. (Little, R.J.A. and D.B. Rubin, Statistical Analysis with Missing Data, Wiley, NY (1987)). Conveniently, the maximum likelihood methods discussed above will produce the correct results, provided that the phenotypes are recorded for all progeny: genotypes for the non-extreme progeny may simply be entered as missing. Using the MAPMAKER-QTL program, the method has been applied to both simulated and experimental data sets.

Decreasing environmental variance via progeny testing: As shown above, the number of progeny needed to map a QTL is proportional to

$$(\sigma_{\text{res}}^2 / \sigma_{\text{exp}}^2) = [(\sigma_G^2 + \sigma_E^2) / \sigma_{\text{exp}}^2] - 1.$$

Typically, the environmental variance exceeds the genetic variance. If σ_E^2 could be reduced, QTL mapping would become considerably more efficient. If the environmental noise results from measurement error, one might either average replicate measurements or try to develop a better assay. More often, environmental noise results from actual physiological differences between genetically identical individuals. In this case, σ_E^2 can be reduced through progeny testing; an individual's phenotype would be taken to be the average phenotype of n of its B2 backcross offspring. The variance of this average is $\sigma_E'^2 = (1/n)[\sigma_G^2 + \sigma_E^2]$, which will be less than σ_E^2 , except for very small n .

-51-

Simultaneous search: Just as environmental noise can be decreased via progeny testing, genetic noise can be reduced by simultaneously studying several intervals containing QTLs. If the genetic variance is large, such an approach may further decrease the number of progeny required. The extension of interval mapping to such simultaneous search; the question of the appropriate LOD score when considering sets of intervals, and the approximate increase in the power of QTL mapping are discussed in the Appendix [A6].

F2 intercrosses and recombinant inbred strains: Although the discussion above concerns the backcross, it applies directly to F2 intercrosses and recombinant inbred strains, with the following modifications:

(i) In an F2 intercross, a QTL with phenotypic effect δ contributes variance $\delta^2/8$ and thus Wright's formula (2) becomes $k = D^2/8\sigma_G^2$. Since F2 intercrosses provide information about twice as many meioses as backcrosses of the same size, fewer progeny are required for detecting QTLs having purely additive effects: only 50-60% as many progeny are needed, depending on the density of the markers used. If a QTL is partly dominant, one of the backcrosses will be more efficient and one less efficient for mapping it. The magnitude of dominance effects can be estimated by explicitly incorporating them into the maximum likelihood analysis via an additional parameter.

(ii) Recombinant inbred strains are analyzed in the same manner as backcrosses, except that the multi-generational breeding scheme that is used to construct recombinant inbred strains increases the effective genetic length of the genome. Compared to a backcross, the density of crossovers is doubled in a

-52-

recombinant inbred produced through selfing and is quadrupled in a recombinant inbred produced by sib mating. Haldane, J.B.S. and C.H. Waddington, Genetics, 16:357-374 (1931). A genetic length of 2G or 4G must be used in place when computing the appropriate LOD threshold, which leads to an increase of about 0.3 or 0.6, respectively, in the threshold required. Although the higher threshold will increase the number of progeny required, the effect is typically offset by the ability to decrease the number of progeny by reducing the environmental variance through replicate phenotypic measurements within each recombinant inbred strain (cf. progeny testing above). Recombinant inbred strains will thus typically be more efficient for QTL mapping than equal number of backcross progeny. However, this advantage may often be negated by the considerable time and effort required to construct large numbers of such strains. However, the ability to replicate phenotype measurements within each recombinant inbred strain leads to a reduction in the environmental variance (cf. progeny testing above). Typically, the latter consideration dominates. A drawback to employing recombinant inbred strains is the considerable time and effort required for their construction.

Although the Results section is mathematical in parts, the Discussion presents the methodology in terms of explicit graphs that allow a geneticist to design crosses to dissect a quantitative trait by using a complete RFLP linkage map.

30 Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation,

-53-

many equivalents to the specific embodiments of the invention described specifically herein. Such equivalents are intended to be encompassed in the scope of the following claims.

-54-

APPENDIX

[A1] To prove Proposition 1, we use the following lemma.

Lemma. Let $x_1, \dots, x_n \geq 0$. For $y \geq 0$, let $s_y = \sum x_i$ and $t_y = \sum x_i^2$, where the sum is taken over the terms $x_i \geq y$. If $t_0/s_0 \geq y$, then

$$s_y \geq \frac{1}{2}[y + \sqrt{y^2 - 4(y s_0 - t_0)}] \quad \text{and} \quad t_y \geq t_0 - y(s_0 - s_y).$$

05 *Proof:* From the definitions and the non-negativity of the x_i , it is clear that

$$s_y^2 \geq t_y \geq t_0 - y(s_0 - s_y).$$

The constraint on s_y then follows by considering the outer terms and applying the quadratic formula. \square

10 In the context of Proposition 1, suppose that the QTLs in the high strain change the phenotype by $x_1, \dots, x_n \geq 0$, respectively. Using the notation above, we have $D = s_0$ and $\sigma_G^2 = t_0/16 = D^2/k$ (because of non-linkage among QTLs and WRIGHT's formula). Taking $y = \epsilon(D/k)$, the result then follows from the lemma since $D_\epsilon = s_y/s_0$, and $V_\epsilon = t_y/t_0$.

15 [A2] Suppose that a QTL lies midway between two flanking markers. Let θ be the recombination fraction between the QTL and either marker and $\psi = 2\theta(1-\theta)$ the recombination fraction between the two markers (ignoring interference). In meioses in which they have not recombined (a proportion $1-\psi$ of the total), the flanking markers act as a single virtual marker linked at recombination fraction γ , where γ is the chance that the QTL recombines with both markers given that the markers themselves have not recombined. By contrast, meioses in which the flanking markers have recombined provide zero information about linkage of the QTL. The ELOD for interval mapping is thus $(1-\psi)$ times the ELOD for a single marker linked at γ which in turn is $(1-2\gamma)^2$ times the ELOD for a marker at 0% recombination. That is,

25
$$\text{ELOD}_{\text{interval mapping}} = (1-\psi)(1-2\gamma)^2 \text{ELOD}.$$

Using the relation $\gamma = \theta^2/[(1-\theta)^2 + \theta^2]$ and simplifying terms, Equation 8a follows.

-55-

[A3] In the idealized *dense-map* case, suppose that markers are available at every point along a chromosome. Suppose that there are no QTLs in the genome. For individual i , the phenotype $\phi_i = N(0,1)$; that is ϕ_i is a random normal variable with mean 0 and variance 1. For individual i , let $x_i(d)$ denote the genotype at a position d cM from the left end of the chromosome ($x_i = 0$ or 1 according to the allele inherited), let $\beta^*(d)$ denote the maximum likelihood estimate of the phenotypic effect of a putative QTL at this position, and let $LOD(d)$ denote the corresponding LOD score. By standard formulas for linear regression,

$$\beta^*(d) = \sum (\phi_i - \bar{\phi})(x_i - \bar{x}) / (\sum (x_i - \bar{x})^2)$$

where $\bar{\phi}$ and \bar{x} are the means of ϕ_i and x_i , respectively. For a large population of size n , the central limit theorem implies that

$$\beta^*(d) = \sum 4\phi_i(x_i - \frac{1}{2}),$$

$$v(d) := \sqrt{n}\beta^*(d) = N(0,1),$$

$$\sigma_{res}^2(d) = \sum (\phi_i - \beta^*(d)x_i(d))^2 \sim n(1 - \beta^*(d))^2,$$

$$\sigma_{exp}^2(d) \sim n[\beta^*(d)]^2 \text{ and}$$

$$LOD(d) \sim \frac{1}{2}(\log_{10} e)(\sigma_{exp}^2(d)/\sigma_{res}^2(d))$$

$$\sim \frac{1}{2}(\log_{10} e)[\sqrt{n}\beta^*(d)]^2.$$

Thus, $LOD(d)$ is asymptotically proportional to the square of a random normal variable $v(d)$ (which incidentally proves that LOD is proportional to χ^2).

Let $r(a,b)$ denote the correlation coefficient for any two random variables a and b . Let d_1 and d_2 denote points on the chromosome and write $d = d_1 - d_2$. From the asymptotic expression for $\beta^*(d)$ above, it follows that

$$r(d) := r(v(d_1), v(d_2)) = \sum \sqrt{n}\phi_i(1-2\theta) \sim 1-2\theta,$$

where θ is the recombination fraction corresponding to the genetic distance $d = |d_1 - d_2|$. Assuming Haldane's map function, $1-2\theta = e^{-2d}$.

To summarize, $v(d)$ is a stationary normal process with covariance function $r(d) = e^{-2d}$. Up to rescaling d by a factor of $\frac{1}{2}$, this is the definition of ORENSTEIN-

-56-

UHLENBECK diffusion and Proposition 2 follows directly (see LEADERBETTER, LINDGREN and ROOTSZEN 1983, Theorem 12.2.9 and discussion following).

05 While only Haldane's map function yields precisely an ORENSTEIN-UHLENBECK diffusion, the proof of Proposition 2 holds in general. The relevant results in LEADERBETTER, LINDGREN and ROOTSZEN (1983) require only that $r(d) \sim 1-2d+o(d^2)$ as $d \rightarrow 0$, which holds for all map functions.

10 [A4] If QTLs with very large effects are segregating, regression analysis is not strictly appropriate (whether in the traditional approach or in the generalization developed in the text) because the phenotypic distribution eventually becomes bimodal. In the extreme, the assumption that the phenotypes are generated by the segregation of a QTL will *always* fit the data much better than the assumption that they follow a normal distribution--even for loci unlinked to any QTLs. Consequently, expressions (3) and (6) fail to give the correct number of progeny required in the case of QTLs with very large effects: indeed, they tend to zero (whereas a minimum positive number of progeny is obviously needed no matter how large the effect of the QTL).
15 To achieve the correct limiting behaviour, the LOD score for a marker at 0 cM can be redefined as the \log_{10} of

$$L(a,b,\sigma^2)/\frac{1}{2}\{L(a,b,\sigma^2)+L(a,-b,\sigma^2)\}$$

20 with $L(a,b,\sigma^2)$ defined in (4). This ratio measures how much more likely the data is to have been generated by a QTL with the hypothesized effect located at the marker locus than by a QTL with this same effect but unlinked to the marker. The ELOD can be found by numerical integration over the distribution for ϕ . In the limit of a QTL with large effect, the expression tends to the traditional LOD score for a qualitative trait used in human genetics.

25 For the QTLs likely to be encountered in practice, this correction is irrelevant. We have used it in computing the number of progeny required in Figures 5 and 6, however, in order that these graphs exhibit the correct limiting behaviour--rather than tending to zero.

30 [A5] For notational convenience, rescale the phenotype so that its mean in the backcross is 0 and encode the two alternative genotypes by the indicator variable $g = -1$ or 1 (rather than 0 or 1, as in the text). Given a true QTL, let $2h$ be the amount by which substituting an allele increases the phenotype and let σ^2 be the residual variance unexplained by the QTL out of the total backcross variance $\Sigma^2 = \sigma^2 + h^2$. Suppose that a marker is located exactly at the QTL. Conditional on the phenotype ϕ of an individual but unconditional on its genotype x at the marker, the
35

-57-

LOD score (comparing the true hypothesis $H_1:(0,b,\sigma^2)$ to the alternative $H_0:(0,0,\Sigma^2)$) is

$$LOD_{\phi} = \sum_{g=0,1} \pi(g|\phi) \log_{10}[z(\phi-bg,\sigma^2)/z(\phi,\Sigma^2)]$$

where $\pi(g=x|\phi)$ is the probability that the individual has marker genotype x given its phenotype ϕ , given by

$$z(\phi-bg,\sigma^2)/[z(\phi-bg,\sigma^2)+z(\phi+bg,\sigma^2)].$$

As claimed in the text, if b is small, LOD_{ϕ} is proportional to ϕ^2 . Now, the probability distribution for ϕ has density

$$p(\phi) = \frac{1}{2}[z(\phi-bg,\sigma^2)+z(\phi+bg,\sigma^2)].$$

Conditional on the phenotype of a backcross progeny deviating from the mean by $\geq L\Sigma$, the LOD score is

$$LOD_{|\phi|\geq L\Sigma} = \int_{|\phi|\geq L\Sigma} (LOD_{\phi})p(\phi)d\phi.$$

Letting $v = b^2/\Sigma^2$ denote the fraction of variance explained by the QTL, straightforward although tedious integration shows that

$$\begin{aligned} S(L) &= LOD_{|\phi|\geq L\Sigma} / LOD_{|\phi|\geq 0} \\ &\approx Q(L) \cdot [1 + 2uLz(L)/Q(L)], \end{aligned} \quad (10)$$

where $u = -v/\log_e(1-v) \approx (1-\frac{1}{2}v)$ and where the approximation in (10) is $o(v^2)$ for small v . For QTLs with small effects, this reduces to Equation (9).

[A6] Interval mapping can be straightforwardly extended to the case of *multiple* intervals explaining a quantitative phenotype: for m intervals, the bracketed term in Equation 7 becomes a sum with 2^m terms corresponding to the possible joint genotypes at the m putative QTLs. Since simultaneous consideration of multiple QTLs reduces the unexplained variance, it may be somewhat easier to detect linkage to the set of loci than to any one individually (cf. LANDER and BOTSTEIN 1986ab). The subtle issue is the appropriate threshold for simultaneous search for m QTLs. In a genome with no QTLs, how high a LOD score might occur by chance? For any particular choice of putative QTLs, the LOD score is

-58-

asymptotically distributed as χ^2 with m degrees of freedom. When considering an entire genome, the LOD score follows a mathematical process known as a χ^2 *random field* (ADLER 1981)--about which somewhat less is known than the ORENSTEIN-UHLENBECK diffusion. Approximate arguments show that the level of the highest excursion of such a χ^2 random field on an entire genome is about $m\pi$ -fold higher than the corresponding level for an ORENSTEIN-UHLENBECK diffusion on the genome. If m QTLs have equal effects, then simultaneous search decreases the number of progeny required to achieve statistical significance by a factor of about $(1-m\sigma^2)/(1-\sigma^2)$, where σ^2 is the fraction of variance explained by each. If the QTLs have unequal effects, it may become possible to detect those with smaller effects by first controlling for those with larger effects. We will discuss simultaneous search for QTLs in more detail elsewhere.

05

10

-59-

CLAIMS

1. A method of mapping genomic regions containing polygenic factors controlling quantitative trait loci comprising the steps of:
 - 05 a. crossing two strains of interest, the strains differing as to at least one trait of interest, to produce progeny;
 - b. carrying out one or more crosses, which are either back-crosses or intercrosses, to produce
10 progeny;
 - c. scoring progeny of the one or more crosses for at least one trait of interest and for selected genetic markers, the genetic markers comprising a genetic linkage map; and
 - 15 d. applying an algorithm designed to maximize the likelihood of the trait of interest to the results of scoring progeny as in step (c).
2. A method of mapping quantitative trait loci in the genome of a higher plant, comprising the steps of:
 - 20 a. crossing two interfertile strains of the higher plant, the two strains differing as to at least one trait of interest, to produce progeny;
 - b. carrying out one or more back-crosses of one of the two interfertile strains and progeny
25 produced in step (a), to produce progeny;
 - c. scoring progeny of the back-crosses for at least one trait of interest and for selected genetic markers, the genetic markers comprising a genetic linkage map; and
 - 30 d. applying an algorithm designed to maximize the

-60-

likelihood of the trait of interest to the
result of scoring progeny as in step (c).

3. A method of genetic dissection of a quantitative trait, comprising the steps of: making an
05 appropriate interspecies back-cross which will detect a specific minimum phenotypic effect and determining the presence or absence of the minimum phenotypic effect.
4. A method of mapping genomic regions containing
10 polygenic factors controlling at least one quantitatively inherited trait of interest, comprising the steps of:
 - a. making a back-cross of two species which differ
15 in at least one quantitatively inherited trait of interest, to produce back-cross progeny;
 - b. scoring back-cross progeny as to the occurrence of selected restriction fragment length polymorphisms;
 - c. constructing a genetic linkage map using the
20 information resulting from step (b); and
 - d. constructing an interval map of quantitative trait loci.
5. A method of designing a cross for genetic dissection
25 of a quantitative trait, comprising the steps of:
 - a. selecting two strains differing for a quantitative trait of interest, the two strains having a phenotypic difference D which is large compared to an environmental variance σ^2_E and a backcross between the two strains having a

-61-

relatively small number of K of QTLs
segregating therein;

- 05 b. specifying a minimum phenotypic effect δ that
the cross of the two strains will be designed
to detect the specific minimum phenotypic
effect δ ensuring that QTLs accounting for a
substantial amount of the phenotypic difference
D will be detected; and
- 10 c. genotyping a number N of the backcross progeny
of the two strains where N is calculated from a
function of spacing between genetic markers in
a map (of ?), threshold T for an LOD score (of
?), and a desired probability β that a false
15 positive occurs; the number N decreasing for
increasing fraction of genetic variances V in
the backcross explained by the QTLs that
account for a substantial amount of the
phenotypic difference D, and the number N
decreasing for smaller K with larger D.

- 20 6. A method of Claim 5 wherein within progeny of a
backcross of one of the strains σ^2_{B1} , variance of
the phenotype is defined by

$$\sigma^2_{B1} = \sigma^2_G + \sigma^2_E$$

where σ^2_G is genetic variance.

- 25 7. A method of Claim 6 wherein K is defined by

$$K = \frac{D^2}{16\sigma^2_G}$$

-62-

8. A method of Claim 5 wherein the step of specifying a minimum phenotypic effect δ includes choosing a value of δ which is from approximately $\frac{1}{2}(D/K)$ to approximately (D/K) .

05 9. A method of Claim 5 wherein the genetic variance V explained by the QTL is defined

$$V = \frac{\delta^2}{16\sigma_{B1}^2}$$

10 where σ_{B1}^2 is the variance of the phenotype in progeny of a backcross of one of the selected strains.

10. A method of Claim 5, further comprising the step of: mapping QTLs when the phenotypic difference D measured in environmental standard deviations is on the order of the number K .

1/15

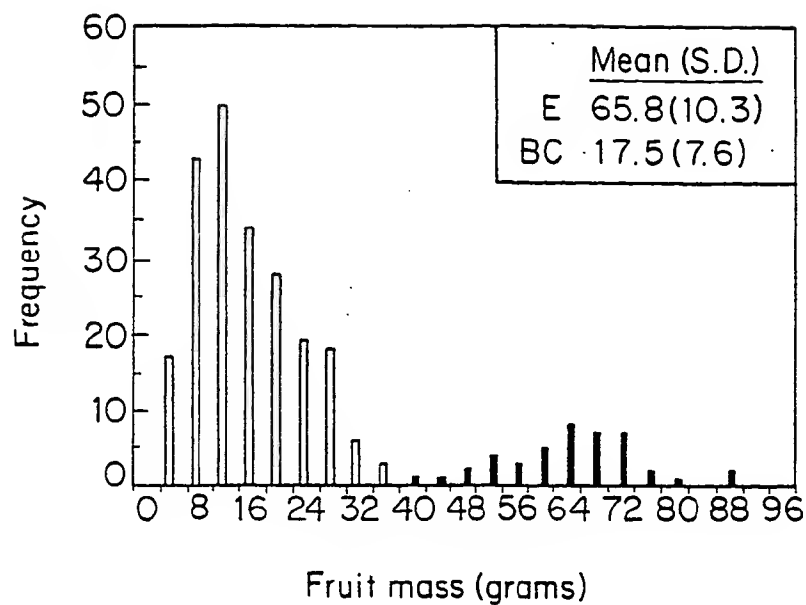


Fig. 1A

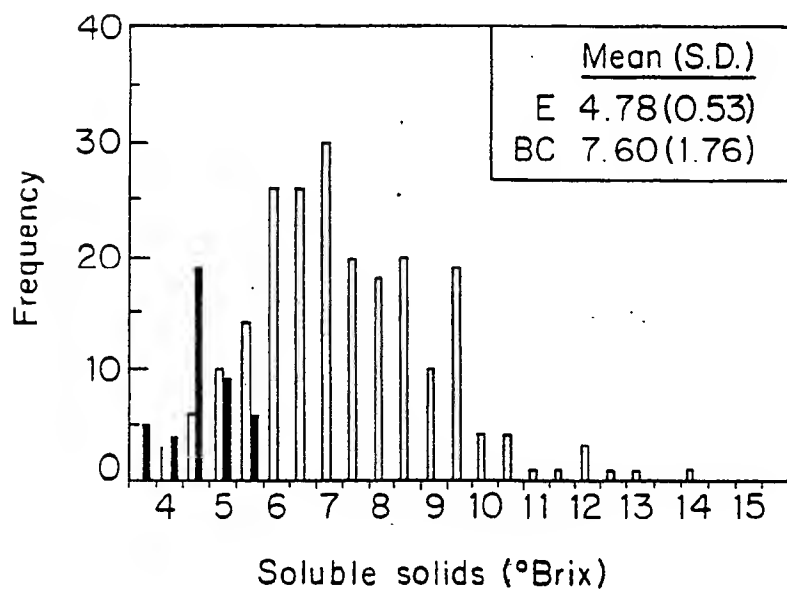


Fig. 1B

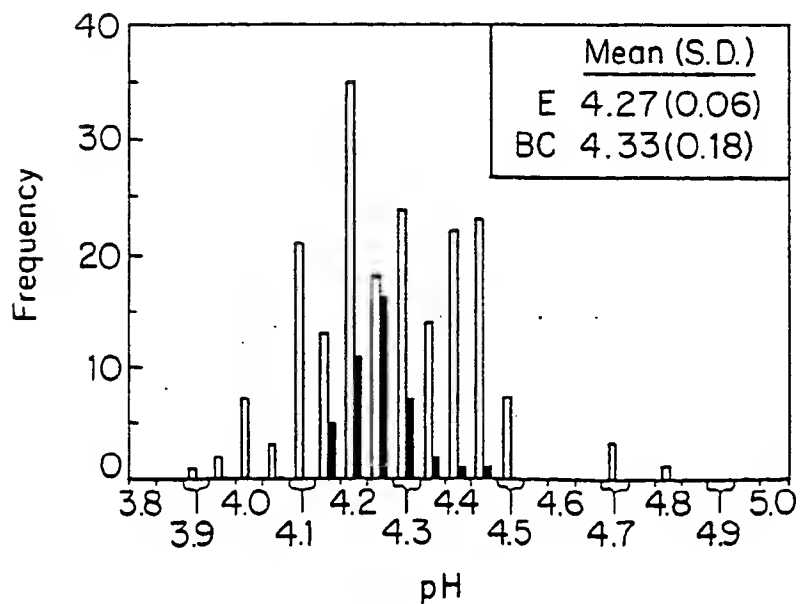


Fig. 1C

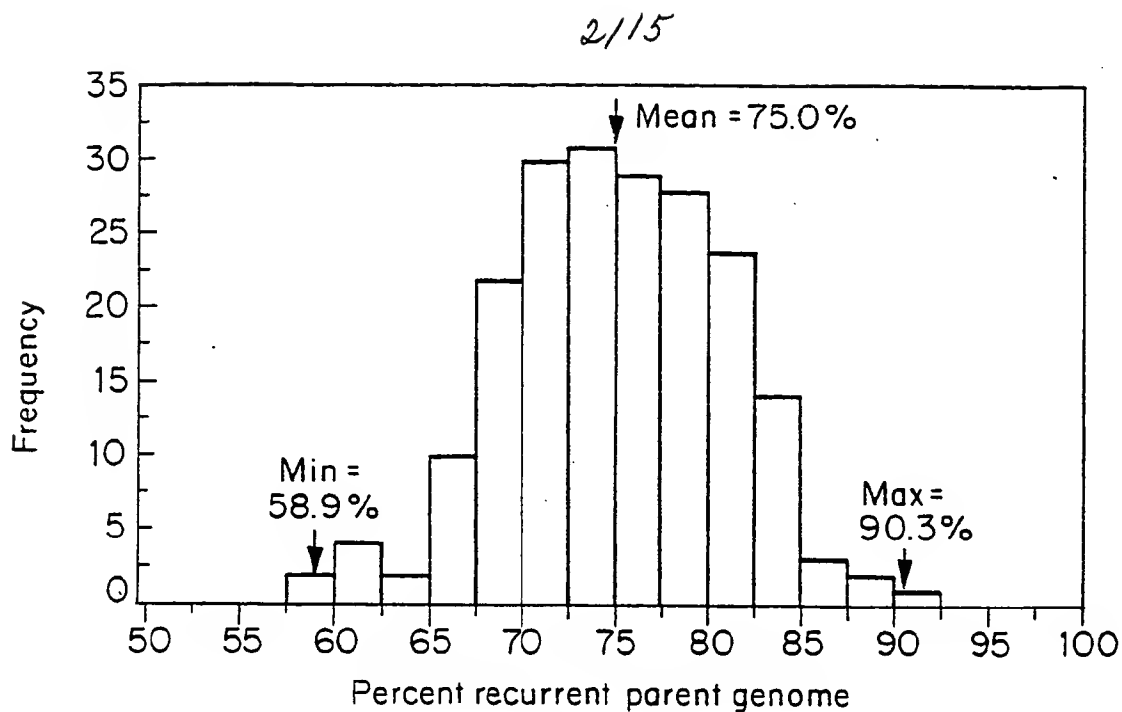


Fig. 2

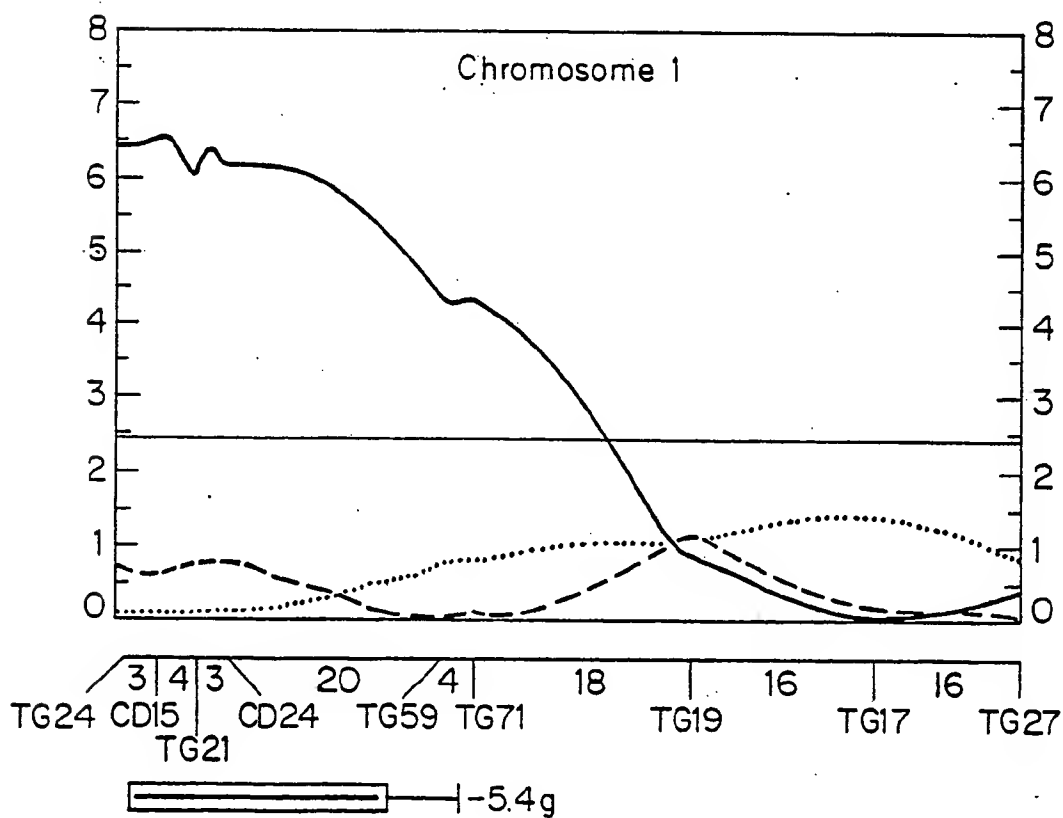


Fig. 3A

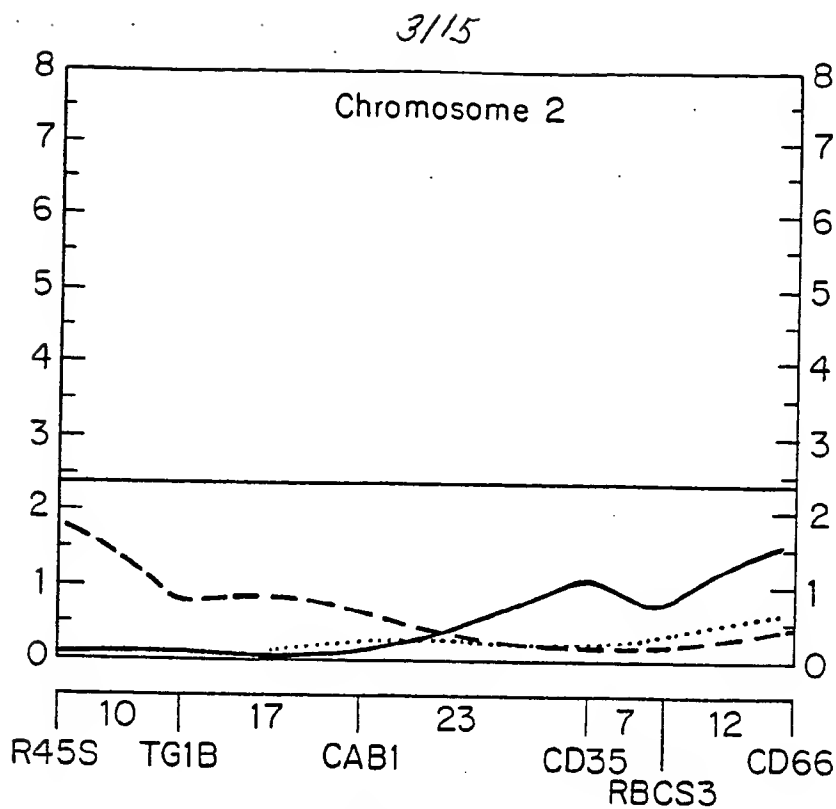


Fig. 3B

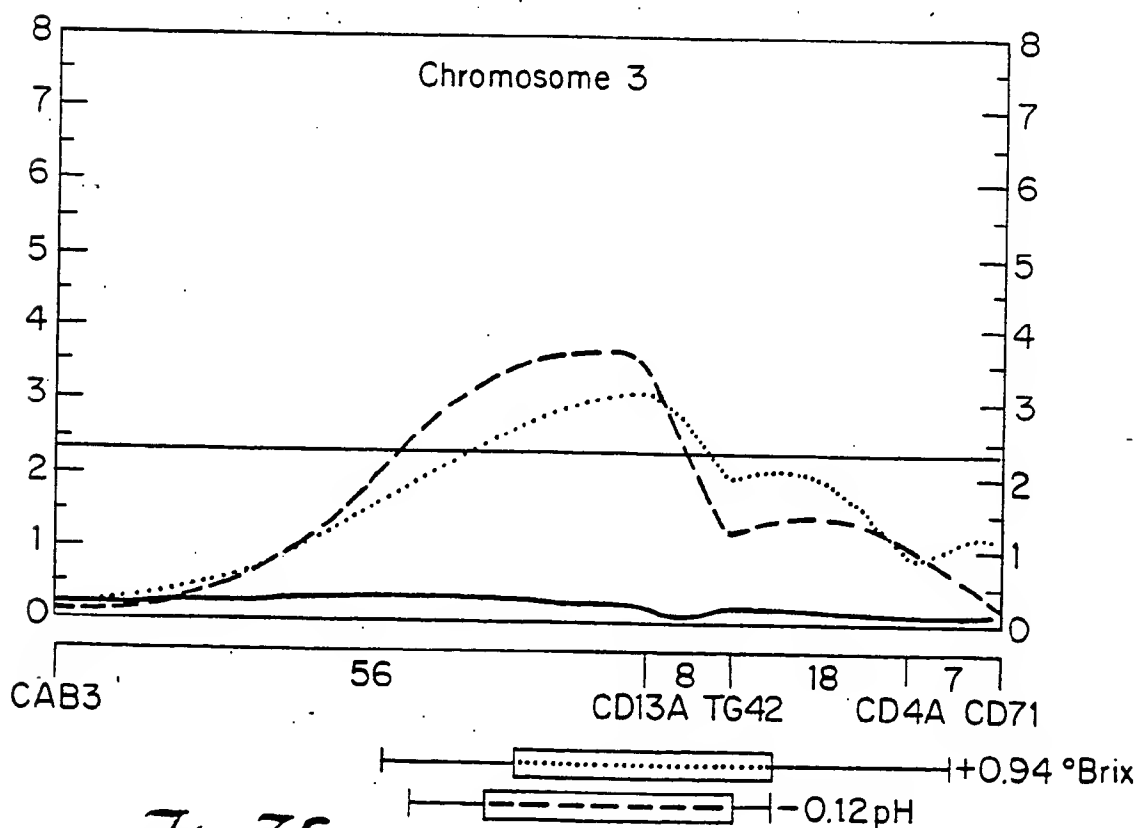


Fig. 3C

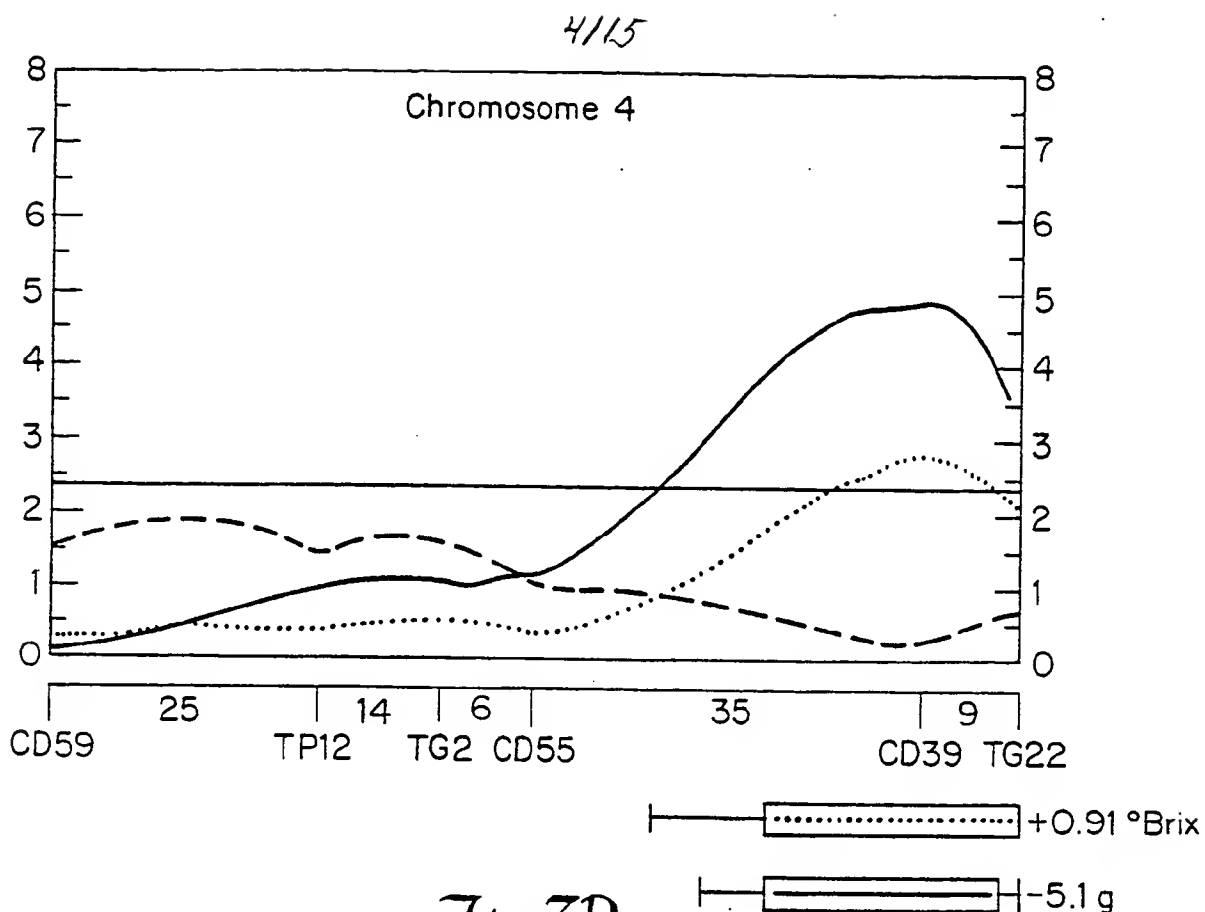


Fig. 3D

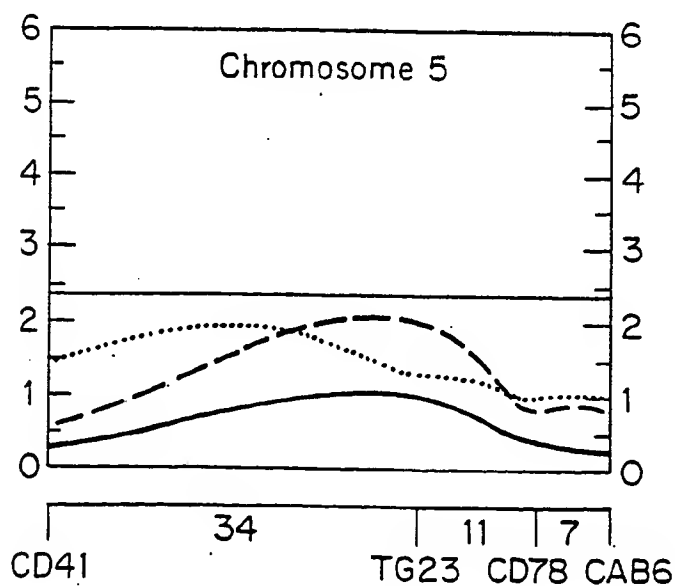
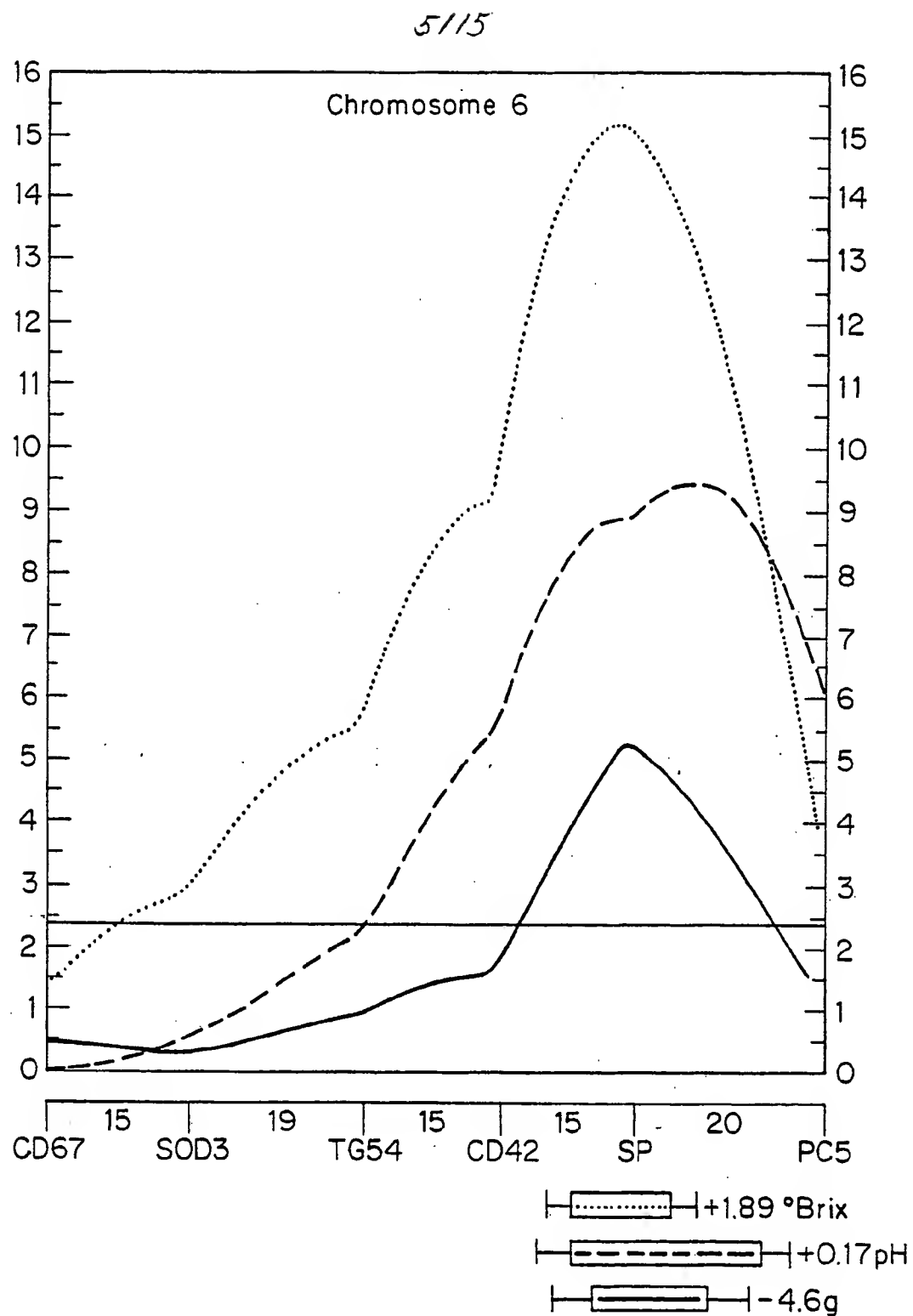


Fig. 3E

*Fig. 3F*

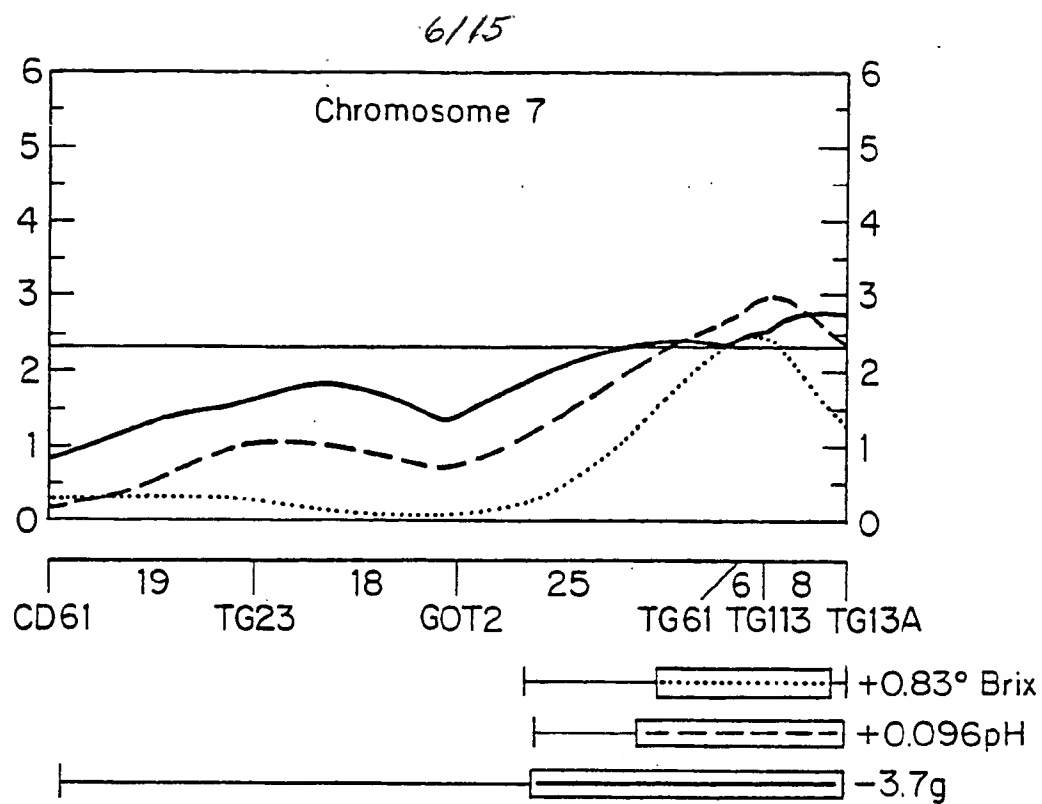


Fig. 3G

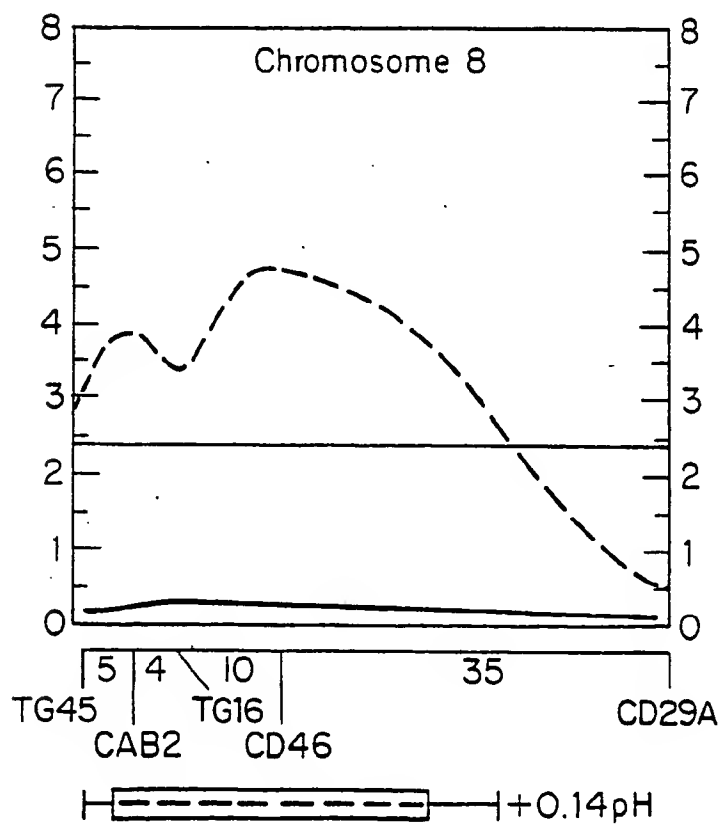


Fig. 3H

7/15

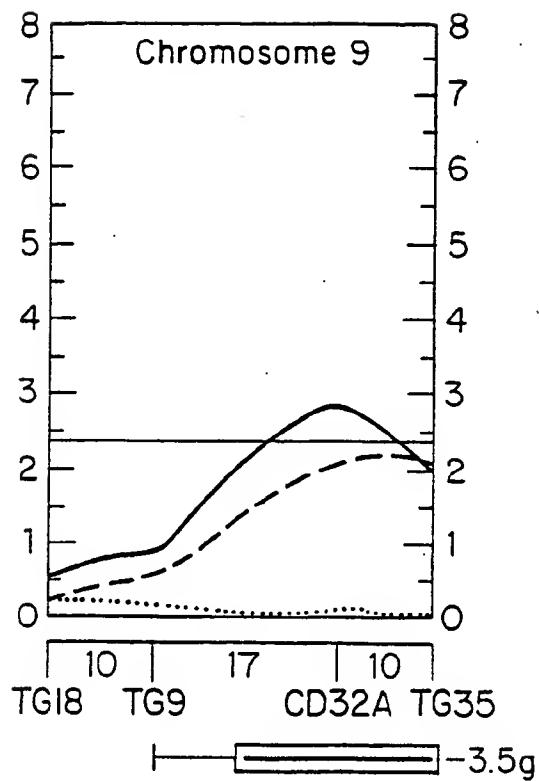


Fig. 3I

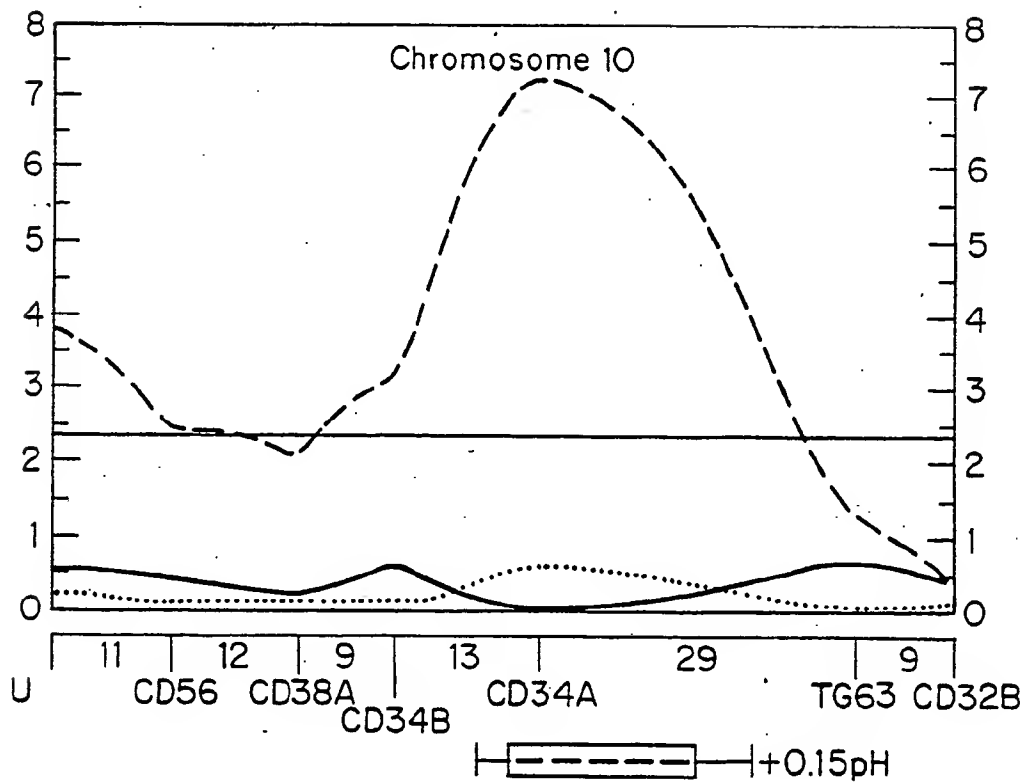


Fig. 3J

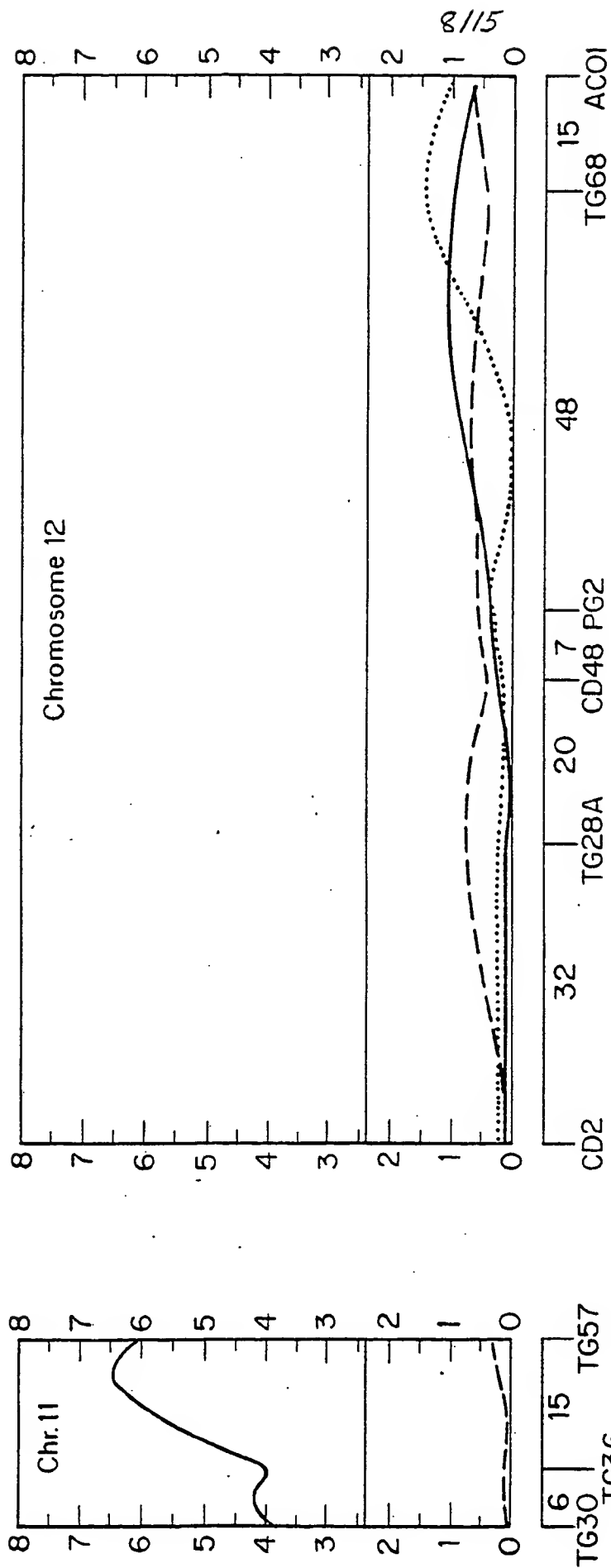


Fig. 3L

Fig. 3K

9/15

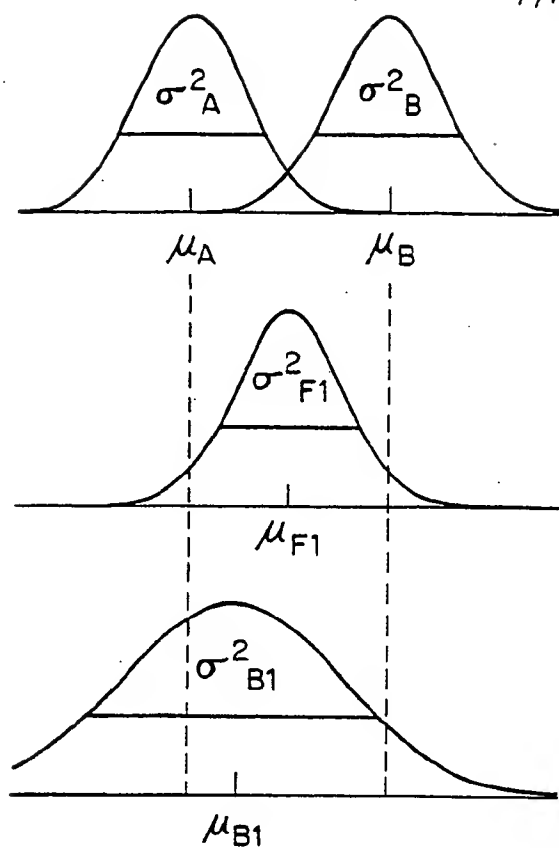


Fig. 4

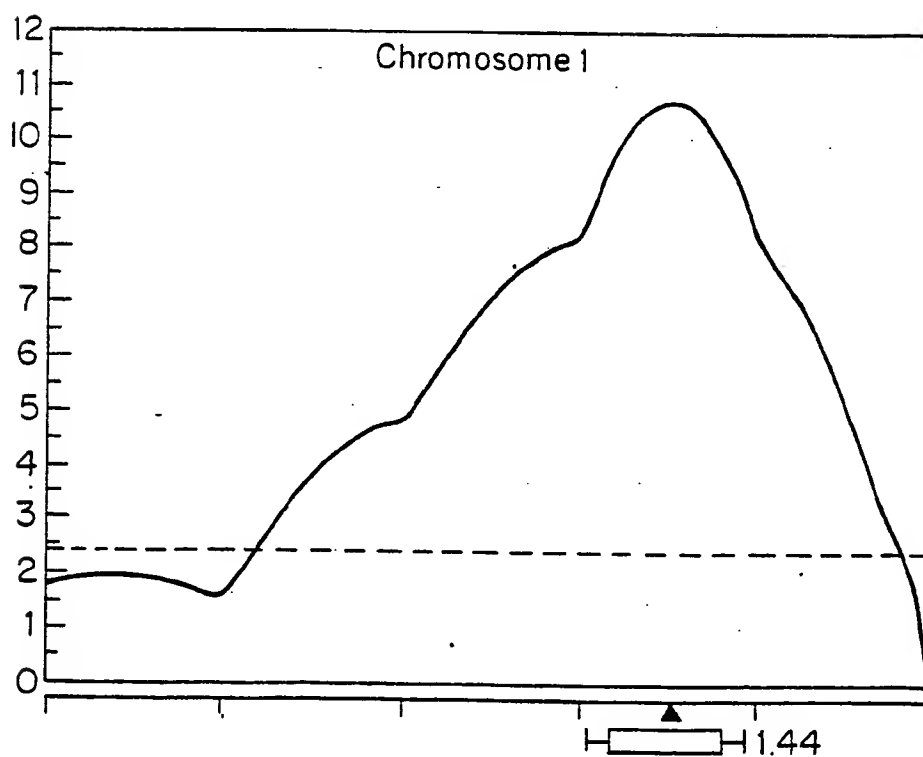


Fig. 5A

10/15

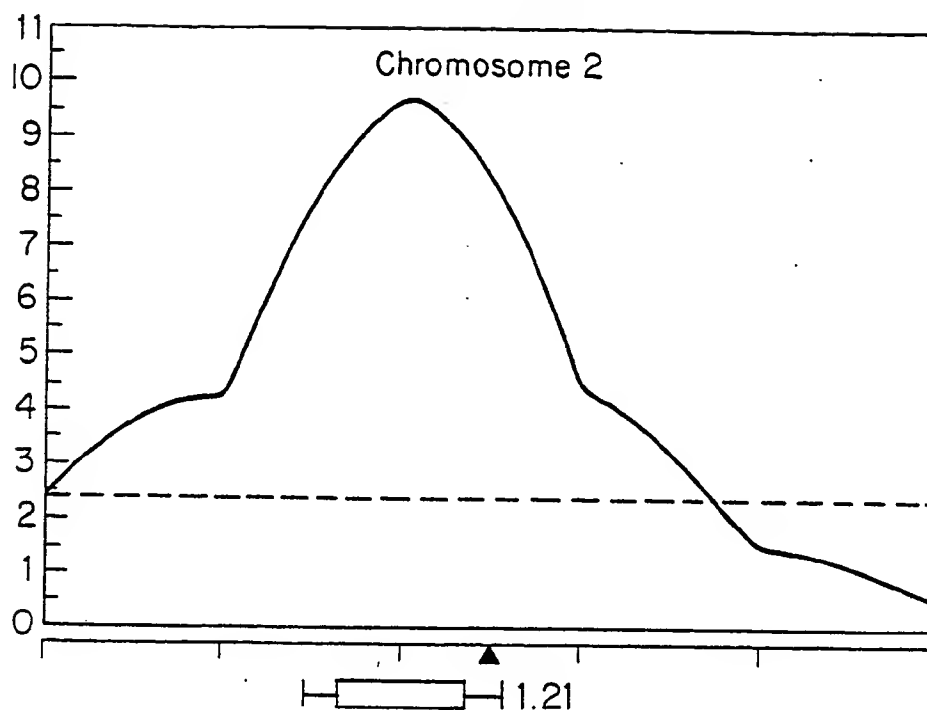


Fig. 5B

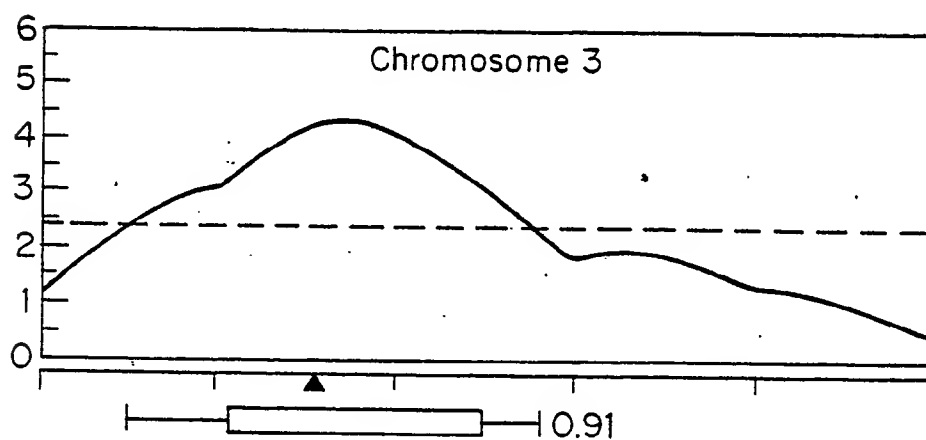


Fig. 5C

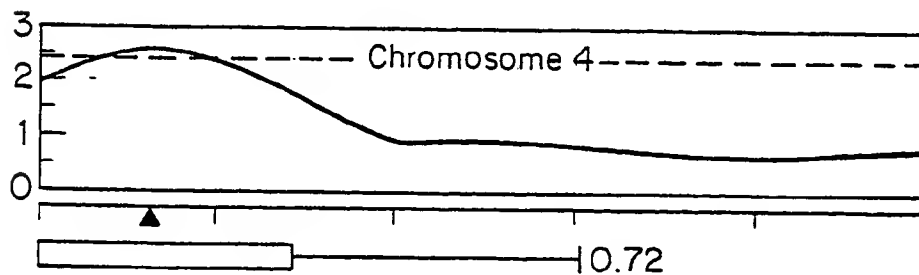


Fig. 5D

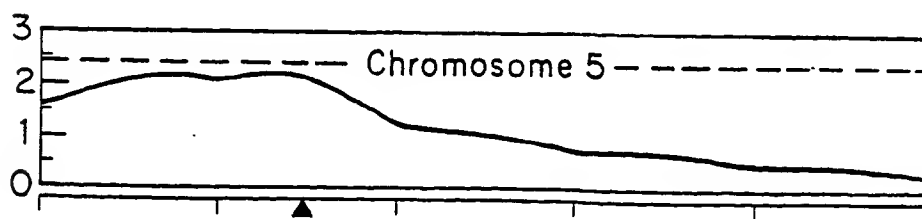
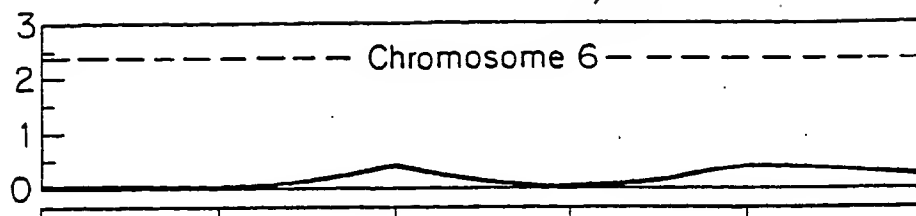
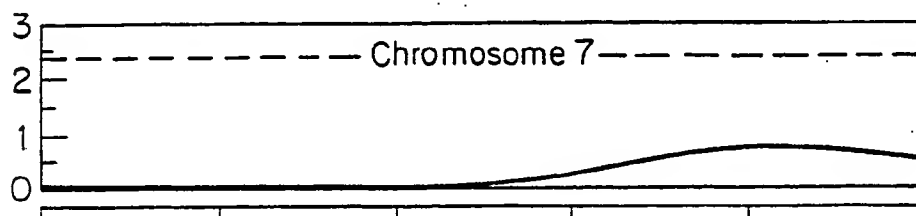
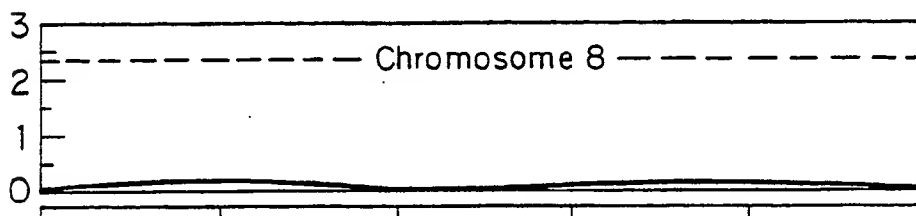
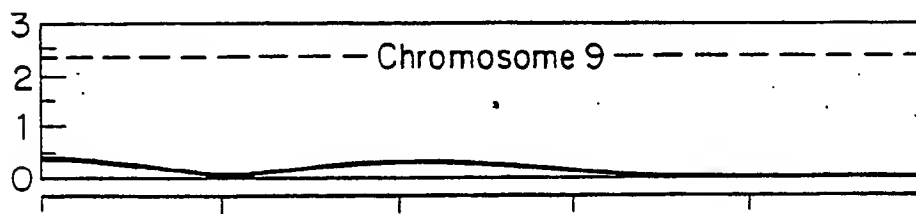
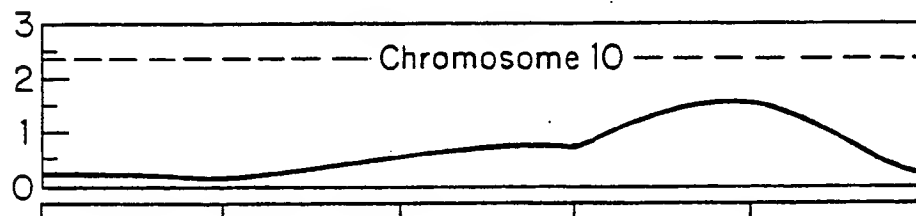
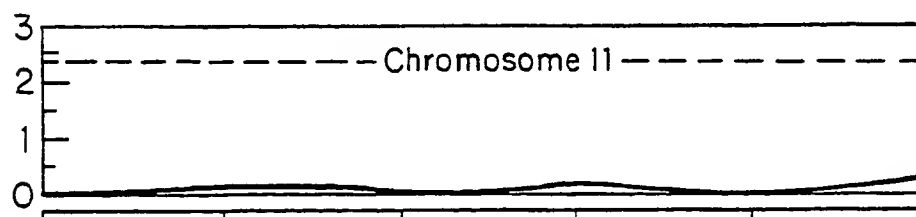
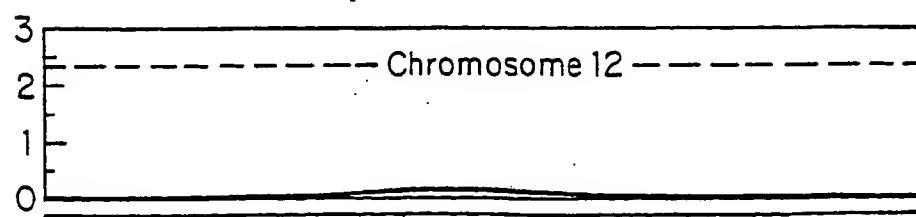


Fig. 5E

11/15

*Fig. 5F**Fig. 5G**Fig. 5H**Fig. 5I**Fig. 5J**Fig. 5K**Fig. 5L*

12/15

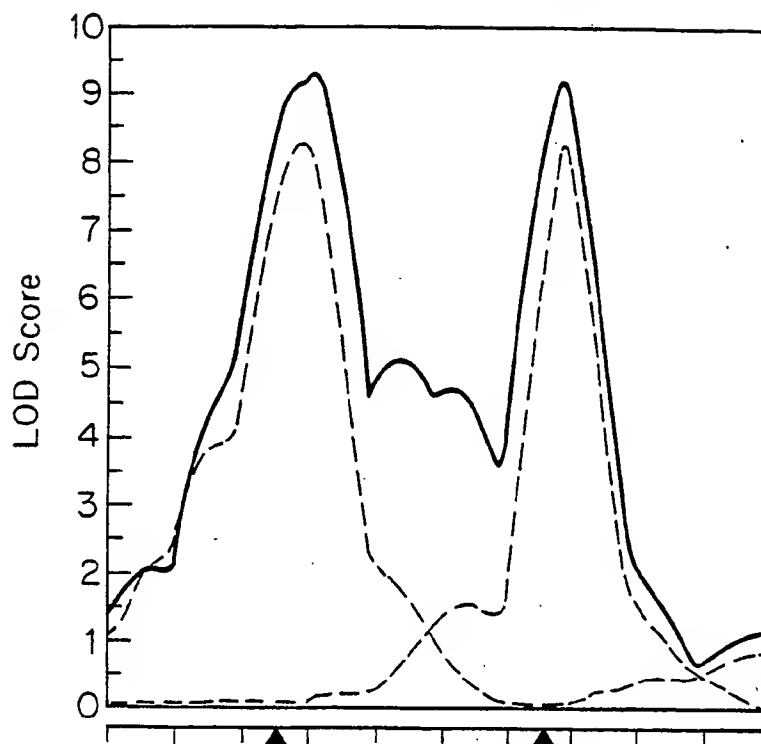


Fig. 6

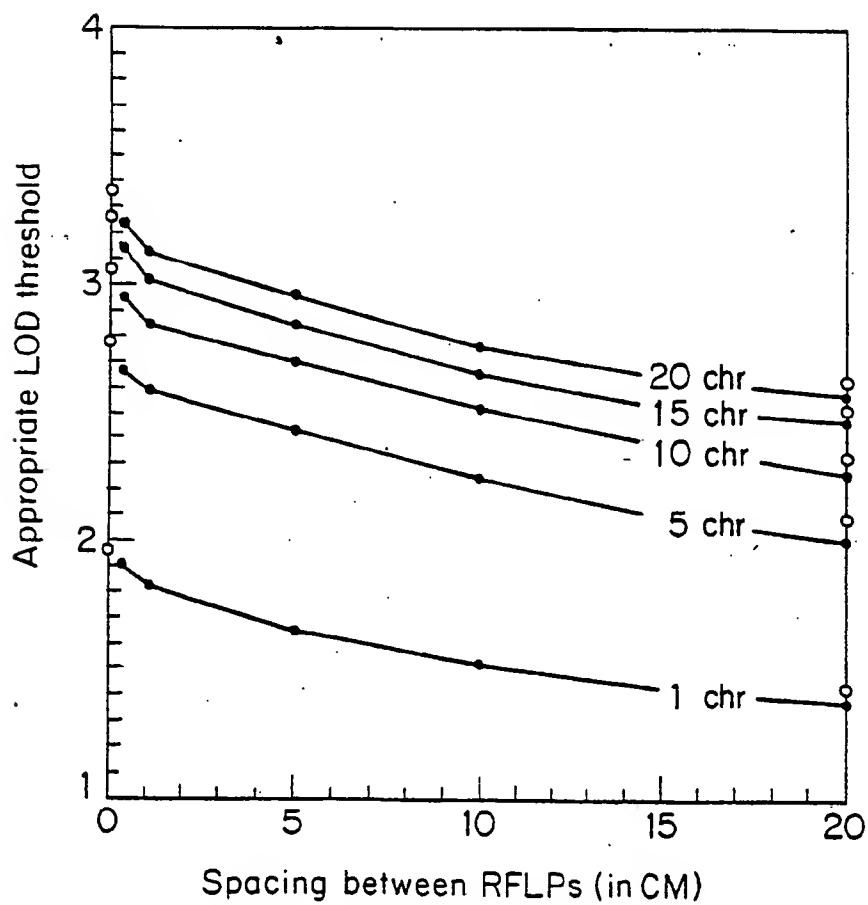


Fig. 7

13/15

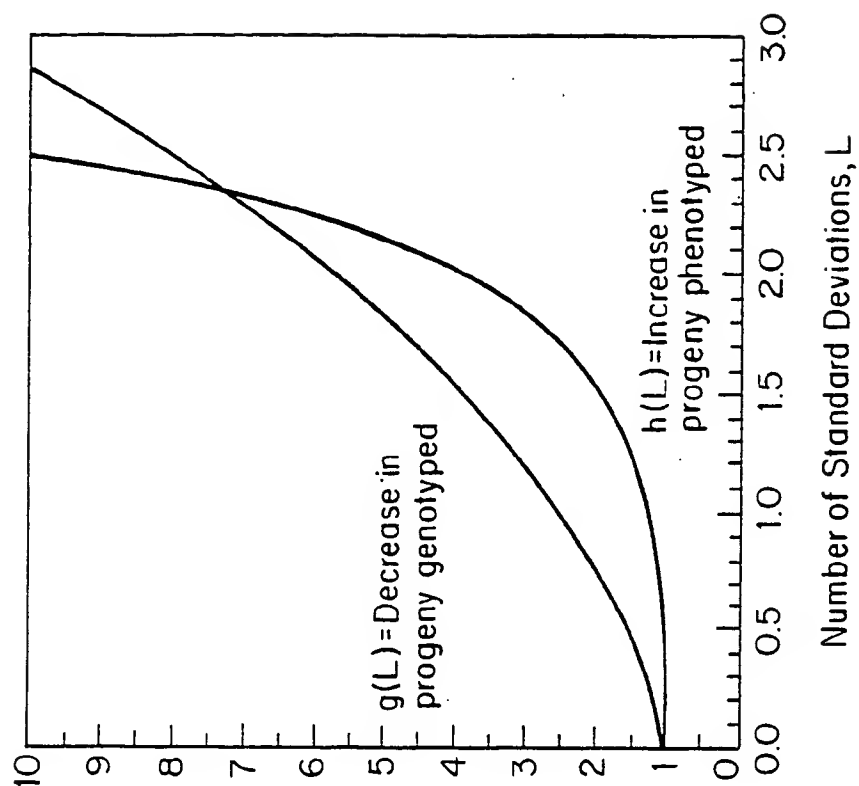


Fig. 9

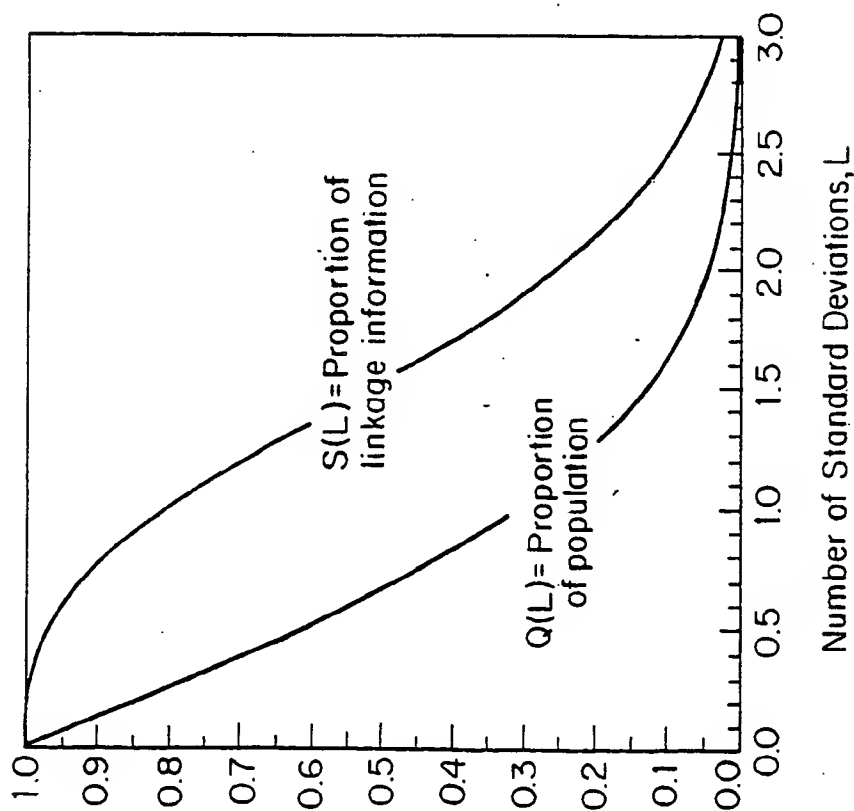
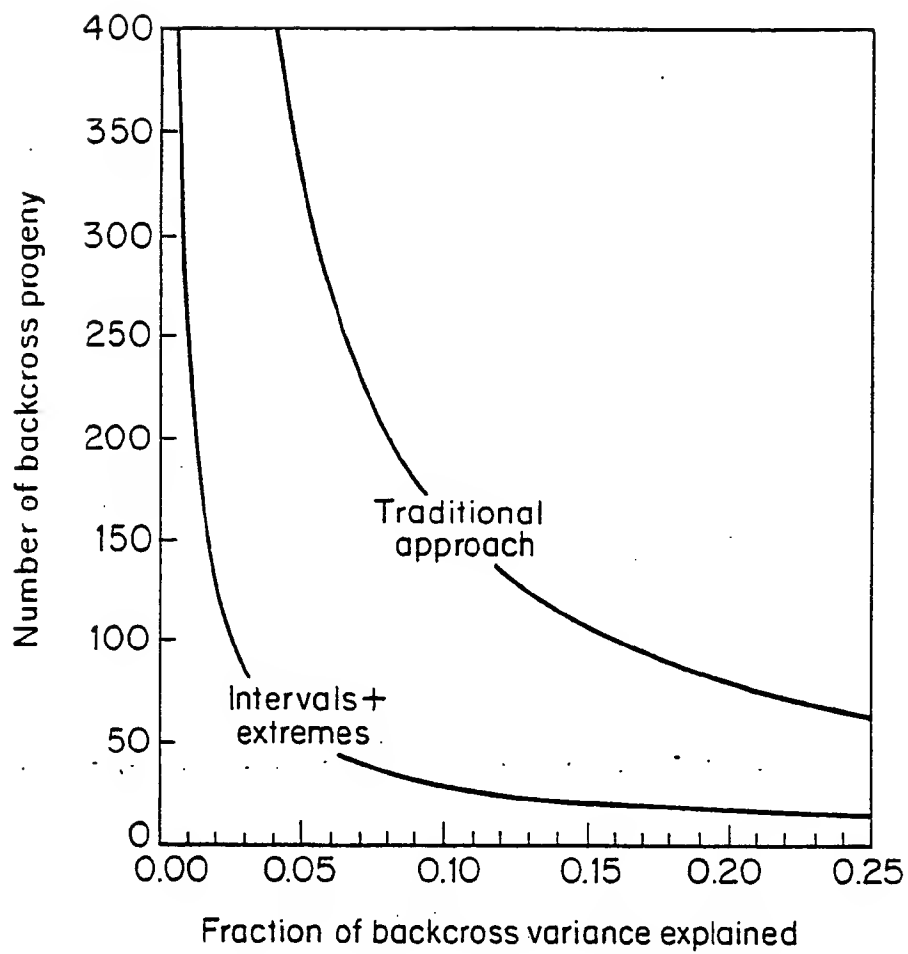


Fig. 8

14/15

*Fig. 10*

15/15

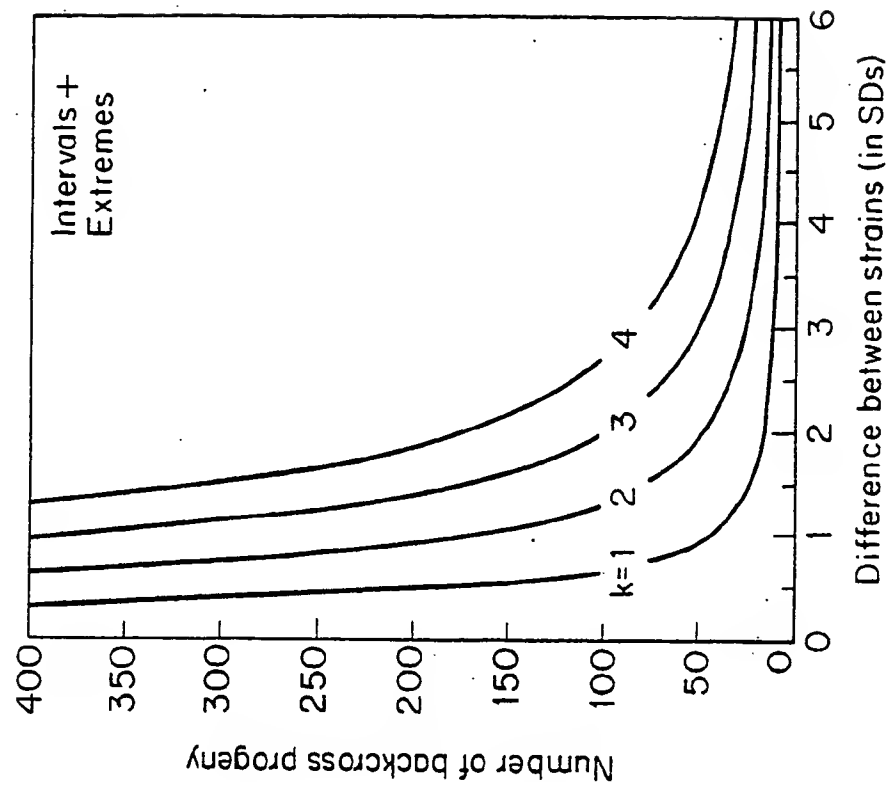


Fig. 11B

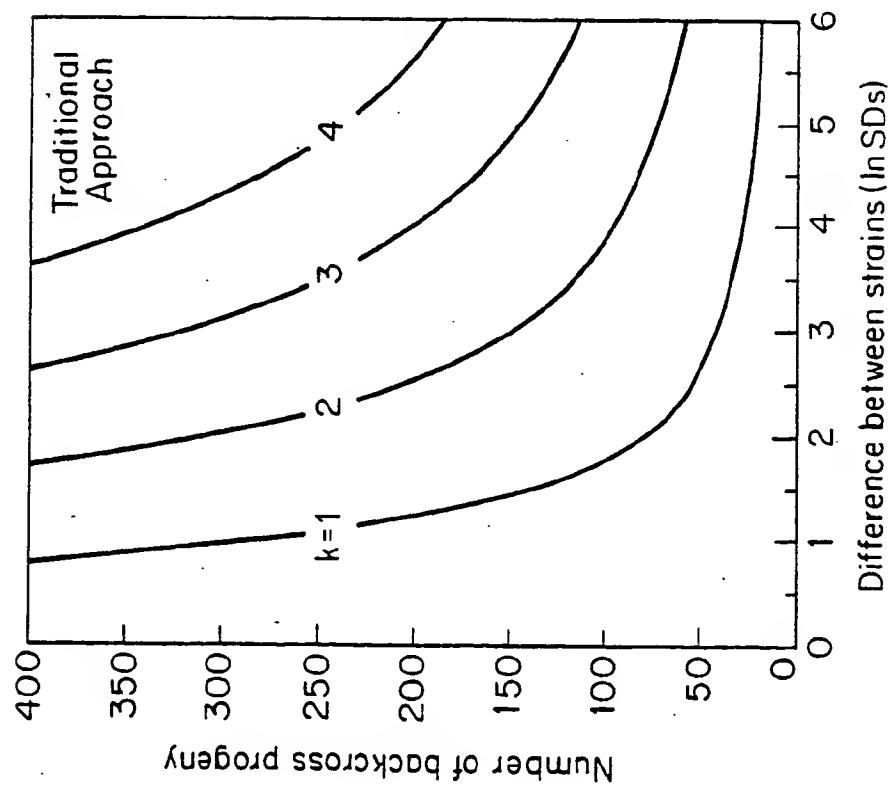


Fig. 11A

INTERNATIONAL SEARCH REPORT

International Application No. **PCT/US 89/04688**

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) *

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC5: C 12 Q 1/68, A 01 H 1/04

II. FIELDS SEARCHED

Minimum Documentation Searched *

Classification System |

Classification Symbols

IPC5

C 12 Q

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched *

III. DOCUMENTS CONSIDERED TO BE RELEVANT *

Category *	Citation of Document, ** with indication, where appropriate, of the relevant passages **	Relevant to Claim No. **
P,X	NATURE, Vol. 335, 1988, Andrew H. Paterson et al: "Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms ", see page 721 - page 726 --	1-10
X	Genetics, Vol. 116, 1987, M.D. Edwards et al: "Molecular-Marker-Facilitated Investigations of Quantitative-Trait Loci in Maize. I. Numbers, Genomic Distribution and Types of Gene Action. ", see page 113 - page 125 see especially page 114 column 2 line 14 - page 117 column 2 line 4 --	1-10

* Special categories of cited documents: **

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the International filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the International filing date but later than the priority date claimed

"T" later document published after the International filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search
14th February 1990

Date of Mailing of this International Search Report

07.03.90

International Searching Authority

EUROPEAN PATENT OFFICE

Signature of Authorized Officer

T. K. WILKES

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
X	BIOMETRICS, Vol. 42, 1986, J.I. Weller: "Maximum Likelihood Techniques for the Mapping and Analysis of Quantitative Trait Loci with the Aid of Genetic Markers ", see page 627 - page 640 see especially page 628 line 22 - page 633 line 19 --	1-10
P,A	WO, A1, 89/07647 (PIONEER HI-BRED INTERNATIONAL, INC.) 24 August 1989, see the whole document -- -----	1-10

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

V. ☒ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND partly unsearchable:¹

This International search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☒ Claim numbers 1-10, because they relate to subject matter not required to be searched by this Authority, namely:

Essentially biological processes for the production of plants as well as mathematical methods, c.f. PCT rule 39.1

2. ☐ Claim numbers..... because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim numbers..... because they are dependent claims and are not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☐ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING :

This International Searching Authority found multiple inventions in this international application as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.
2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:
3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:
4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

- ☐ The additional search fees were accompanied by applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

SA 32472

08/11/89

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A1- 89/07647	24/08/89	NONE	

